Technical Report:

Diagnostic analysis using data from the National Health Services (NHS)

Background

The NHS is focused on optimising healthcare infrastructure to meet growing population needs. Analysing utilisation trends is crucial for informed decision making on budget allocation, capacity expansion, or maximising existing resources. Stakeholder opinions on budget allocation vary, promoting the need for data exploration and recommendations. With a wealth of information and numerous questions, distilling concerns into four concise problem statements is essential for effective data exploration.

Problem statements:

1. There is a lack of clarity on patient engagement in practices across England.
2. There is uncertainty about the proportion of patients actually attending their appointments.
3. There is ambiguity surrounding appointment types, the type of professionals available, and service settings.
4. There is no clear strategy for using external sources (X) to find popular hashtags, which may help improve public health communication and increase patient engagement.

In addition to the questions posed by the NHS, additional business questions were identified from the problem statements:

	Business questions
1)	How many practices are there in England?
2)	What is the number of attended appointments?

3)	What is the number unattended appointments?
4)	What is the number of appointments where patient's attendance was unknown?
5)	What is the distribution of appointments across different actual duration intervals?
6)	What is the count of attended appointments, and which appointment mode was utilised for these appointments?
7)	What is the count of unattended appointments, and which appointment mode was utilised for these appointments?
8)	How does the attendance of appointments vary across different healthcare professional types, and what is the distribution of attended appointments among these types?
9)	What is the distribution of attended appointments across various healthcare professionals and appointment modes?

Analytical approach, patterns and discussion

The data files, appointment_durations.csv, appointments_regional.csv and national_categories.xlsx were imported into Python using pandas and numpy libraries. DataFrames 'ad', 'ar', and 'nc' were created to represent the respective datasets.

Methods for used DataFrame exploration:

- df.shape: returns the DataFrame dimensions.
- df.columns: returns column labels.
- df.dtypes: returns data types.
- df.head( ): displays the first 5 rows of the DataFrame.
- df,isnull( ): returns Boolean expression, 'True' or 'False' indicating missing values.
- df.info( ): provides summary of memory usage, data types and non-null values for each column.
- df.describe( ): provides descriptive statistics.

- Appointment month and date information was converted into the Datetime format, to optimise calculations and improve efficiency in handling date-related information.

Business question:

1) **How many practices are there in England?**
2) **What is the number of attended, unattended and unknown appointments?**

To address the above business questions, the number of practices in England's 'sub_icb_location_name' was determined by utilising the .value_count() method on relevant columns in the 'nc' and 'ar' DataFrames. This analysis revealed a total of 106 practice locations **(see Figure 1).** For attended, unattended and unknown appointments, the 'appointment_status' column in the 'ar' DataFrame was analysed **(see figure 2, 3 & 4).** Clear and concise outputs were presented using print statements and f-strings, revealing specific counts for each appointment status.

	sub_icb_location_name	number_of_locations
0	NHS North West London ICB - W2U3Z	13007
1	NHS Kent and Medway ICB - 91Q	12637
2	NHS Devon ICB - 15N	12526
3	NHS Hampshire and Isle Of Wight ICB - D9Y0V	12171
4	NHS North East London ICB - A3A8R	11837
...	...	...
101	NHS North East and North Cumbria ICB - 00N	4210
102	NHS Lancashire and South Cumbria ICB - 02G	4169
103	NHS Cheshire and Merseyside ICB - 01V	3496
104	NHS Cheshire and Merseyside ICB - 01T	3242
105	NHS Greater Manchester ICB - 00V	2170

106 rows × 2 columns

Figure 1: Output showing that there are 106 practice locations in England.

```
# Determine the number of attended appointments
attended_appointments = ar['appointment_status'].value_counts()['Attended']

# Using a print statement with a docstring
print(f"The number of attended appointments is: {attended_appointments}")
```

```
The number of attended appointments is: 232137
```

Figure 2: Code and output showing the number of attended appointments as 232137.

```
# Determine the number of unattended appointments
unattended_appointments = ar['appointment_status'].value_counts()['DNA']

# View the ouput
print(f"The number of unattended appointments is: {unattended_appointments}")
```

```
The number of unattended appointments is: 163360
```

Figure 3: Code and output showing the number of unattended appointments as 163360.

```
# Determine the number of appointments where patient attendence is unknown
unknown_appointments = ar['appointment_status'].value_counts()['Unknown']

# View the ouput
print(f"The number of appointments where patient attendance is unknown: {unknown_appointments}")
```

```
The number of appointments where patient attendance is unknown: 201324
```

Figure 4: Code and output showing the number of unknown appointments as 201324.

Interpretation:

Even though, patients mostly attended their appointments instances where appointment attendance was unknown or unattended was considerably high. According to NHS England

(2019) each year over 15 million general practice appointments are missed due to patients' failure to show without notifying the surgeries. Out of 307 million schedules appointments with healthcare professions, e.g. therapists, other practice staff and GP's, 5% are missed and this results in approximately 15.4 million vacant sots. With each appointment averaging a cost of £30 the total expense for the NHS exceeds £216 million, and this is without considering the disruption to other patients and staff. While unattended and unknown appointments **(Figure 3 & 4)** appear lower than attended, these figures would still impact the NHS' total expenses. Furthermore, the 'unknown' data poses a significant challenge for meaningful interpretations, as it is not known whether these appointments were attended or not.

Business question:

3) What is the count of attended/unattended appointments, and which appointment mode was utilised for these appointments?

To analyse attended appointments, unattended appointments and service setting, the approach involved creating DataFrames ('appointment_mode_attended' **(see Figure 5)** and 'appointment_mode_not_attended **(see Figure 6)** using the 'ar' DataFrame.

- The .groupby ( ) method was applied to group data by 'appointment_mode' and .size( ) calculated counts.

```python
# Create a DataFrame that shows the count of attended appointments grouped by the 'appointment_mode'.
appoinment_mode_attended = ar[ar['appointment_status'] == 'Attended'].groupby('appointment_mode')\
.size().reset_index(name='attended_appointments')

appoinment_mode_attended
```

	appointment_mode	attended_appointments
0	Face-to-Face	64478
1	Home Visit	48608
2	Telephone	60796
3	Unknown	29526
4	Video/Online	28729

Figure 5: Code and output showing the distribution of attended appointments by appointment mode.

```python
# Create a DataFrame that shows the count of unattended appointments grouped by the 'appointment_mode'.
appoinment_mode_not_attended = ar[ar['appointment_status'] == 'DNA'].groupby('appointment_mode')\
.size().reset_index(name='unattended_appointments')

# View the ouput
appoinment_mode_not_attended
```

	appointment_mode	unattended_appointments
0	Face-to-Face	53770
1	Home Visit	21386
2	Telephone	49696
3	Unknown	23784
4	Video/Online	14724

Figure 6: Code and output showing the distribution of unattended appointments by appointment mode.

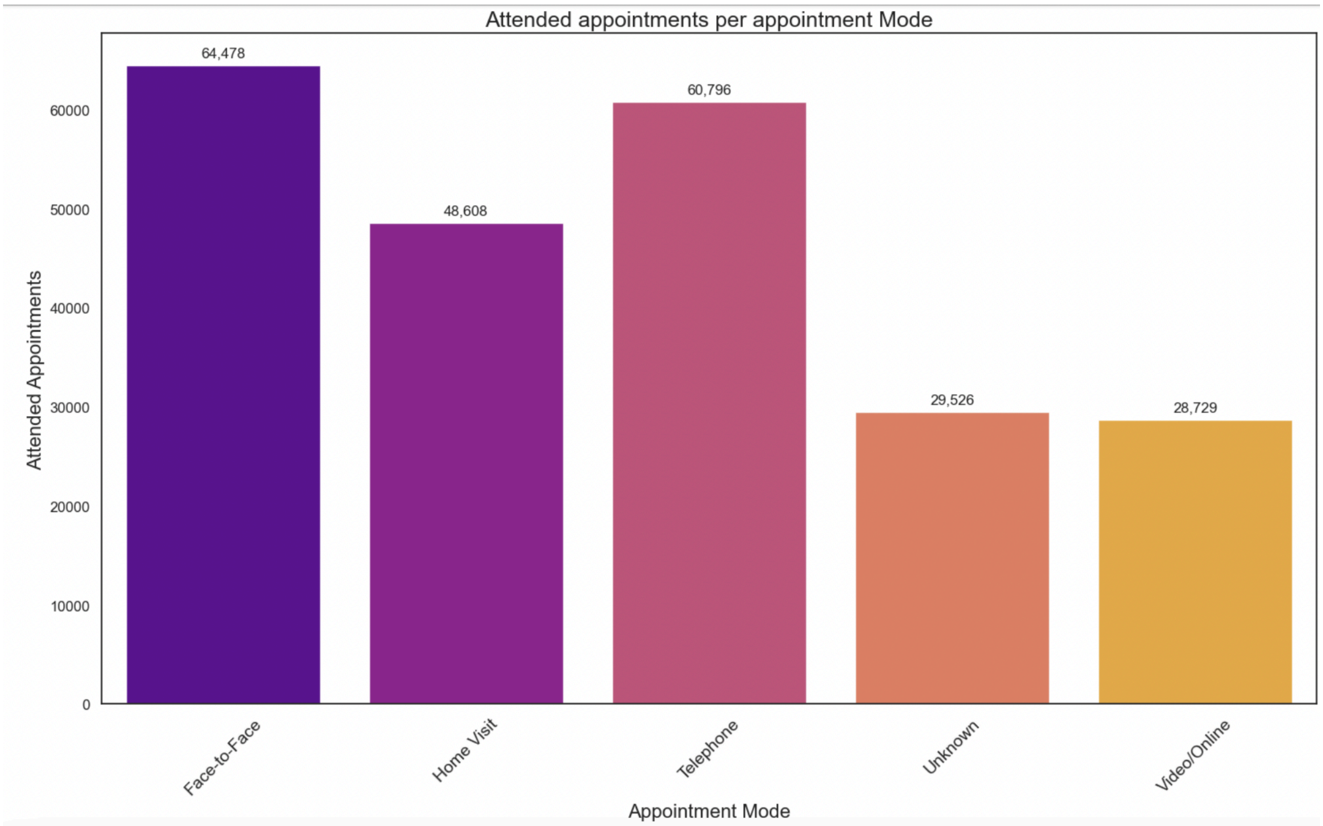

Figure 7: Bar chart showing the distribution of attended appointments by appointment mode.

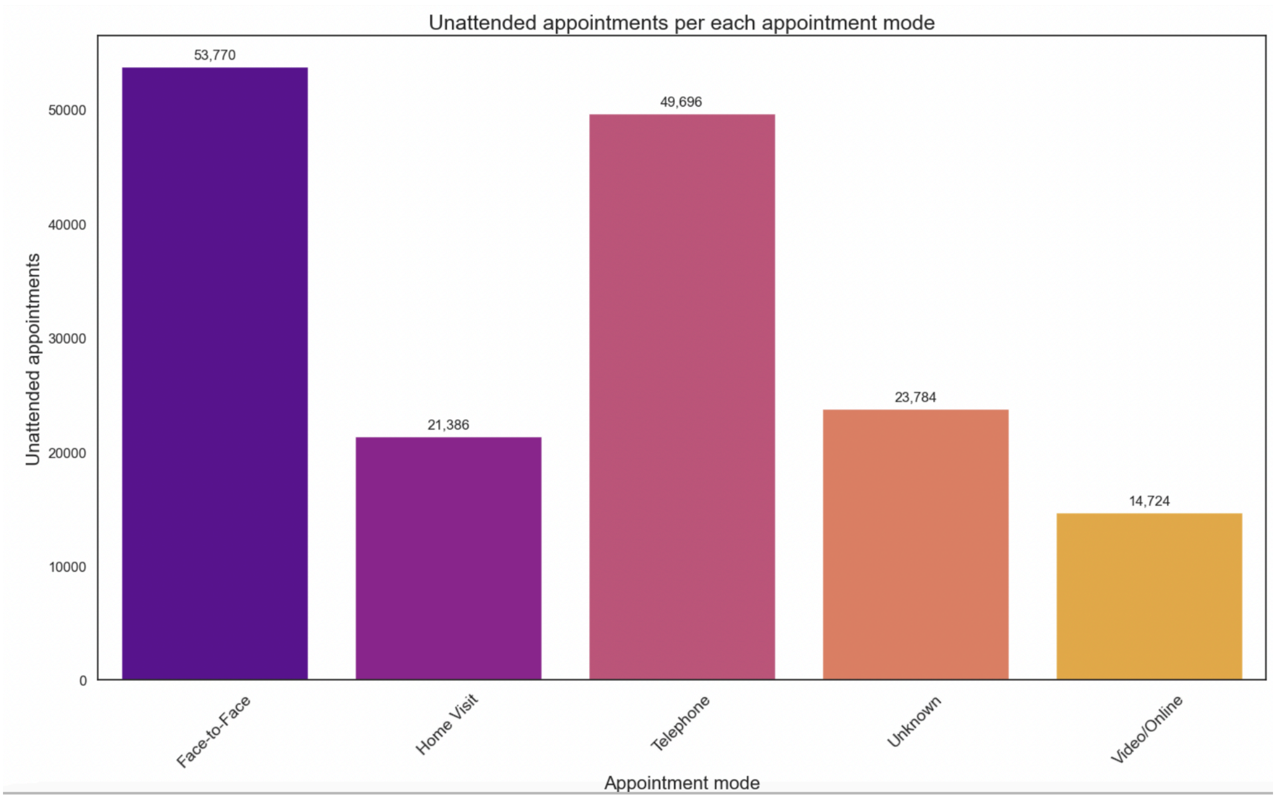

Figure 8: Bar chart showing the distribution of unattended appointments by appointment mode

Interpretation:

Figures 5 and 6 depict the distribution of attended and unattended appointments across appointment modes, revealing patient engagement insights. Examples include 64,478 face-to-face attended and 53,770 unattended appointments. Notably, for attended and unattended appointments unknown appointment modes occurred more frequently than Video/Online appointments **(Figure 7 & 8).**

There are a couple of limitations to discuss in light of this insight:

1. The appointment mode is determined locally by practices, which introduces subjectivity and potential variation in interpretation, leading to different logging practices. For example, a video/online appointment may be classified as face-to-face in one practice and different in another.
2. The inability of practices using the Cegedim GP system to provide data on appointment mode causes a higher proportion of appointments to be categorised as 'Unknown'.

This dual challenge of potential subjectivity/misinterpretation and system-specific data increases uncertainty in the accuracy of appointment mode date, necessitating caution in interpretation and drawing definitive conclusions.

Business question:

4) **How does the attendance of appointments vary across different healthcare professional types, and what is the distribution of attended appointments among these types?**

Interpretation:

While GP's were actively involved in 87,868 attended appointments, 53,417 appointments fell under the 'Unknown' category. Exploring health care professional distribution and appointments **(Figure 9)** brought attention to inherent limitations in the HCP type data. Misreporting instances since October 2017 have led to potential misclassifications of healthcare professionals. Notably, only GP data is consistently collected, resulting in categorisation of all other types as 'Other Practice Staff'. The 'Unknown' category arising from missing data might be attributed to inaccuracies in staff input for healthcare professional types, introducing potential misclassification such as GP appointments being placed under 'Other Practice Staff'. This highlights challenges in the methods used to capture healthcare professional information.

```python
# Create a DataFrame that shows the count of attended appointments grouped by 'hcp_type'
appoinment_attended_hcp = ar[ar['appointment_status'] == 'Attended'].groupby('hcp_type')\
    .size().reset_index(name='attended_appointments') .sort_values(by='attended_appointments', ascending=False)

# View the resulting DataFrame
appoinment_attended_hcp
```

	hcp_type	attended_appointments
1	Other Practice staff	90852
0	GP	87868
2	Unknown	53417

Figure 9: *Code and output showing the distribution of health care professionals and attended appointments.*

Visualisation and insights

Visualisations were created to indicate the number of appointments for each service setting per season (summer, autumn, winter and spring).

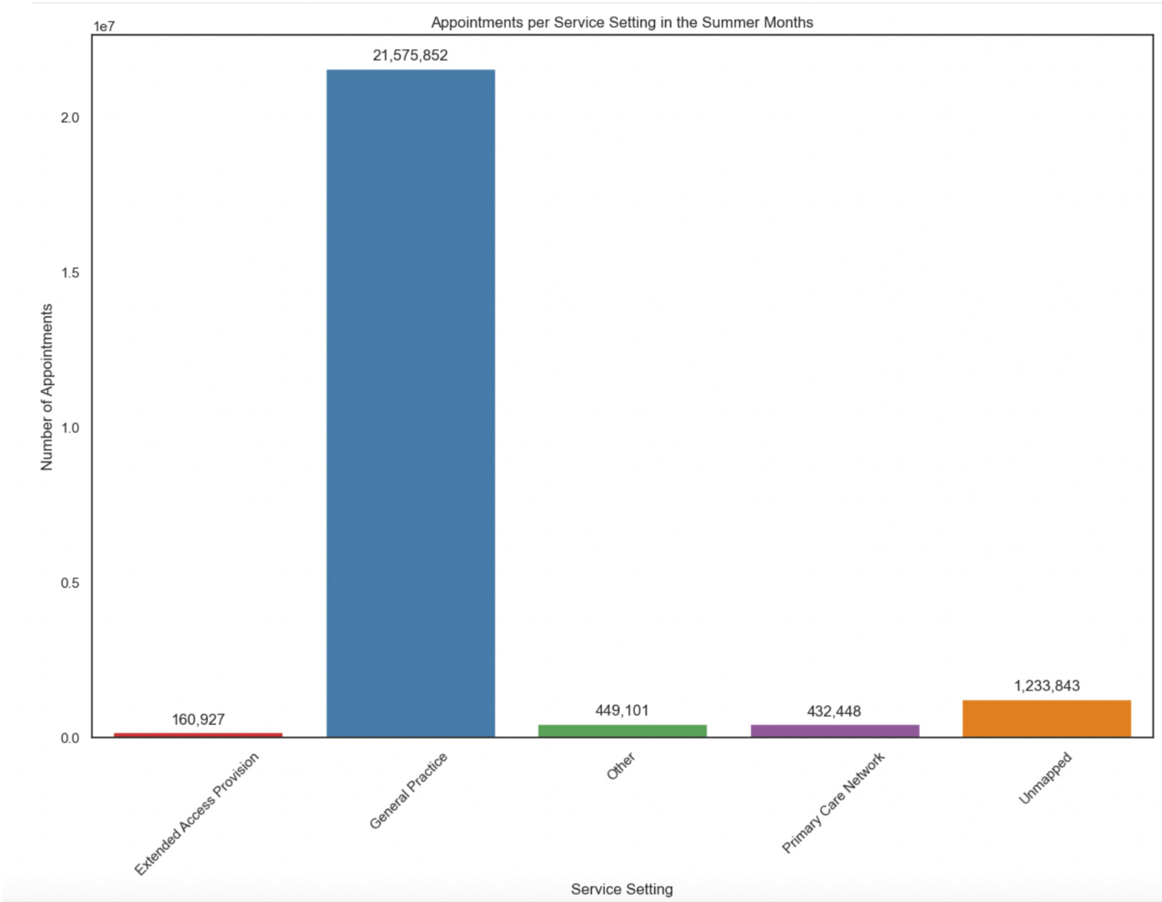

Figure 10: *Bar chart plotting summer months.*

- A bar chart **(Figure 10)** was chosen over a line plot for summer months (June, July, August) as data points aggregated in August causing a lack of data spread across the specified months

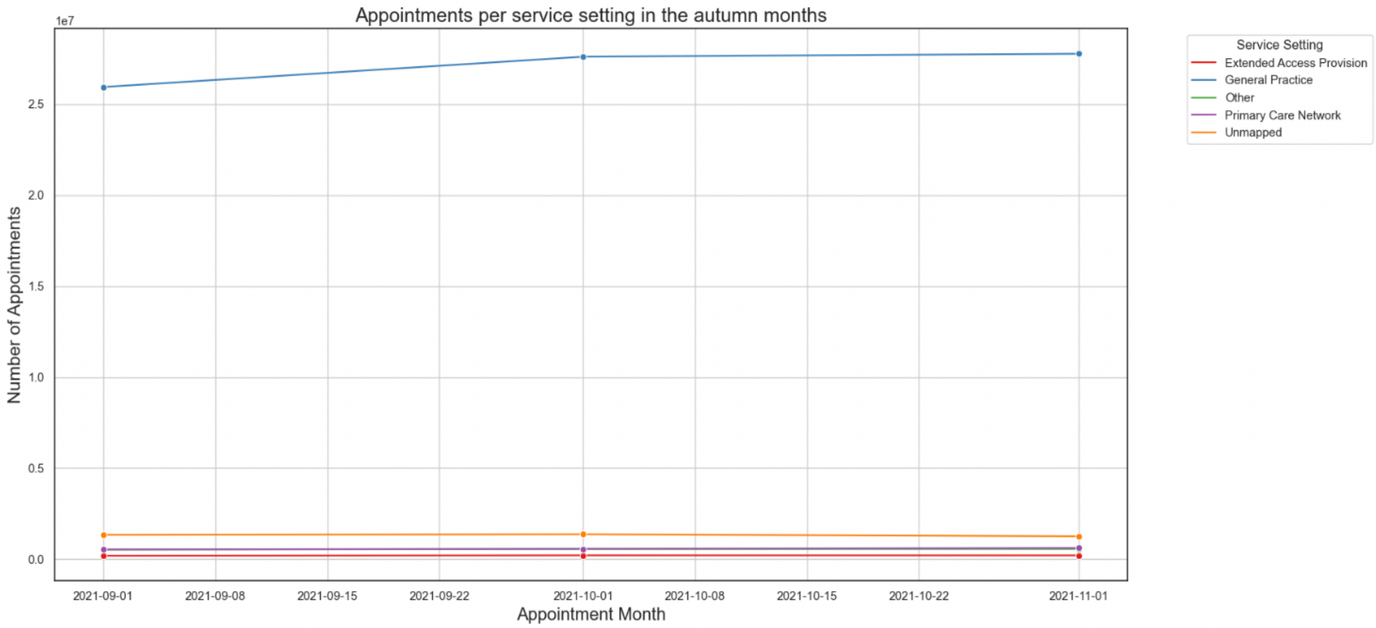

Figure 11: *Line chart plotting autumn months.*

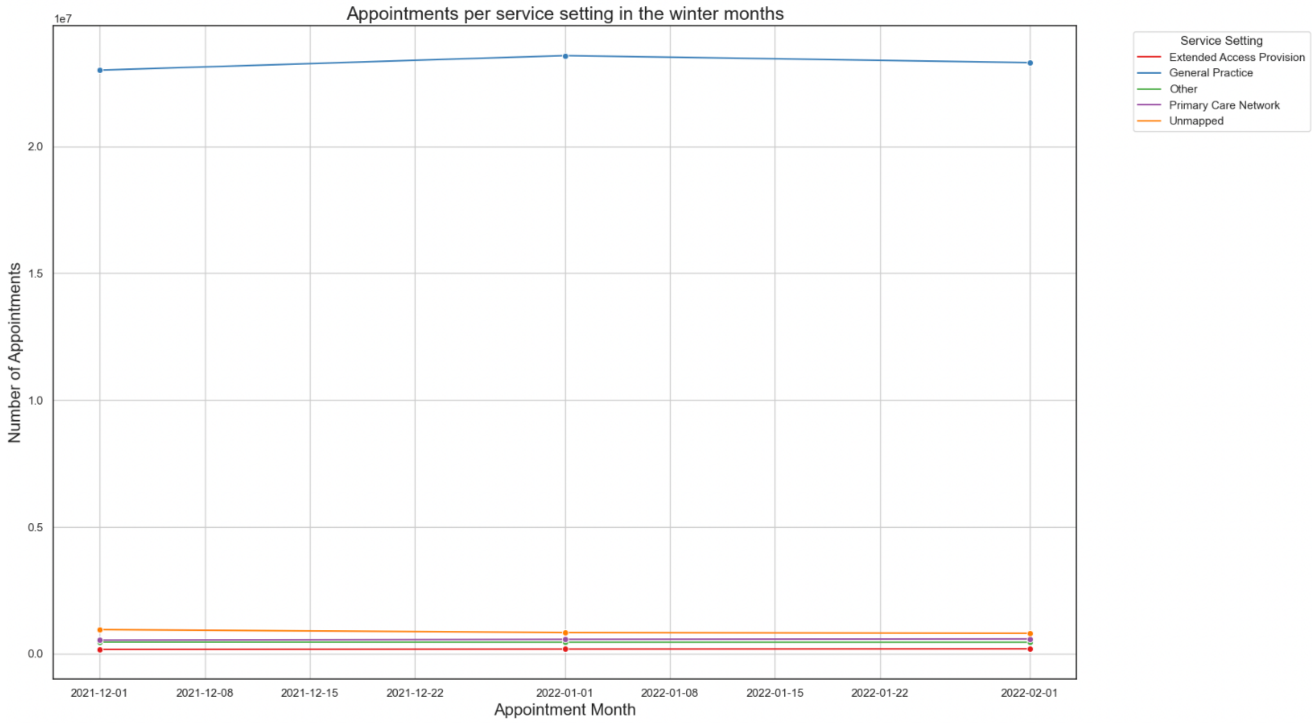

Figure 12: *Line chart plotting winter months.*

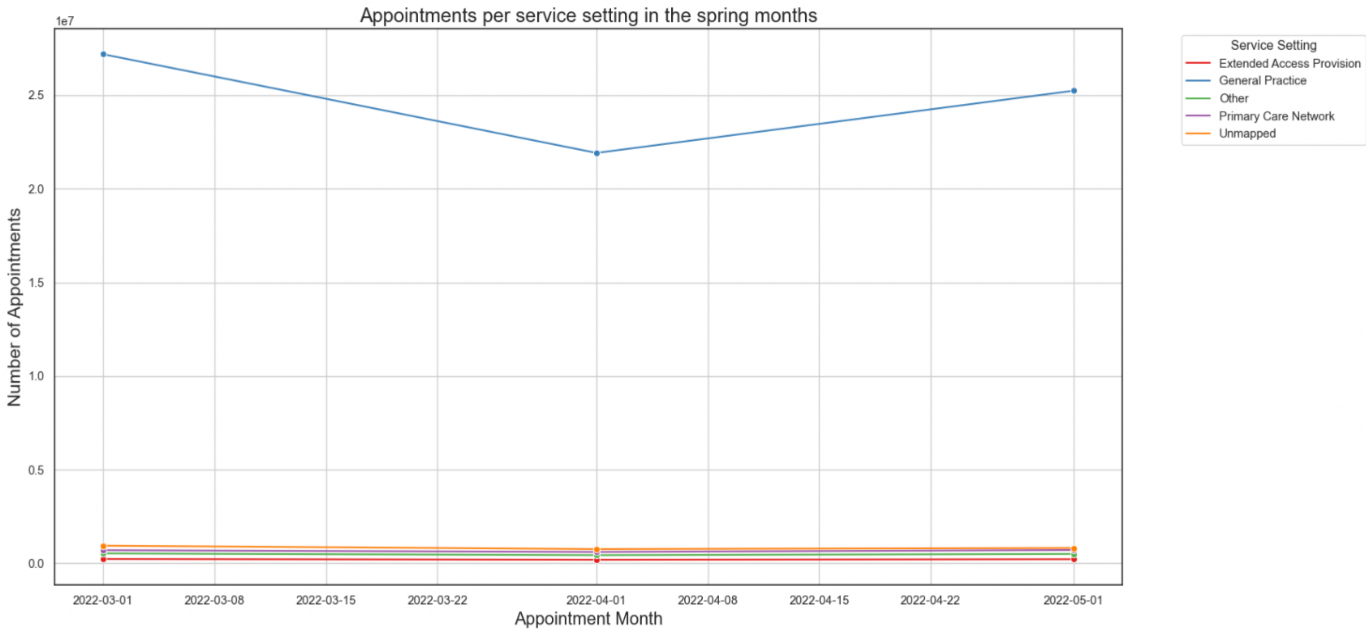

Figure 13: *Line chart plotting spring months.*

Interpretation:

- General practice consistently dominates total appointments per month across all seasons with other service settings maintaining relatively low appointment volumes.
- Autumn general practice appointments exhibit a steady rise in appointment.
- Winter general practice appointments peak in January, gradually declining thereafter.
- Spring general practice appointments peak in early March, followed by a rapid decline until April, after which appointments begin to increase again.

The use of line charts here are beneficial for visualising the temporal trends of appointments across different service settings during different seasons. It enables a clear depiction of changes in appointment over time, making it easier to identify patterns and fluctuations in appointments for each service setting.

Limitation:

- There was issue with Y-axis scaling as a result of Seaborn's automatic scaling constraints and the sheer largeness of the data being explored, which impacts the visualisation effectiveness.

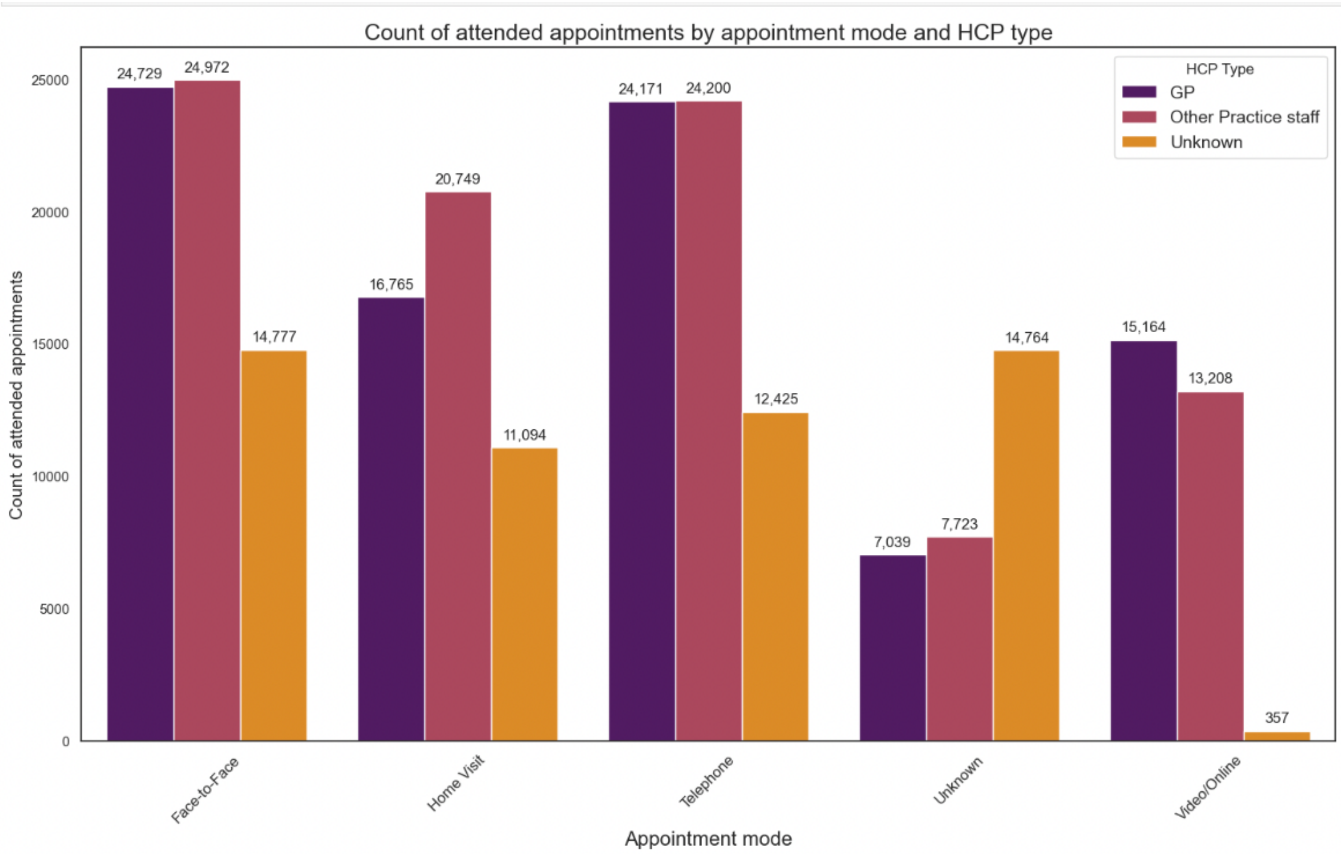

Figure 14: *Bar chart showing how attended appointments are distributed across different appointment modes and healthcare professional types.*

Interpretation:

- Primarily, face-to-face and telephone appointment modes witnessed the highest attendance.
- 'Other Practice Staff' emerged as the dominant professional type for face-to-face consultations and Home Visits.

- Both ‘Other Practice Staff’ and GPs demonstrated similar proportions of attendance in telephone appointments.

The use of a bar chart is suitable for this visualisation, as it effectively compares the distribution of appointments across different appointment modes and HCP types. The use of colour (hue) enhances the differentiation between HCP types and appointment mode to enhance readability and comprehension.

.

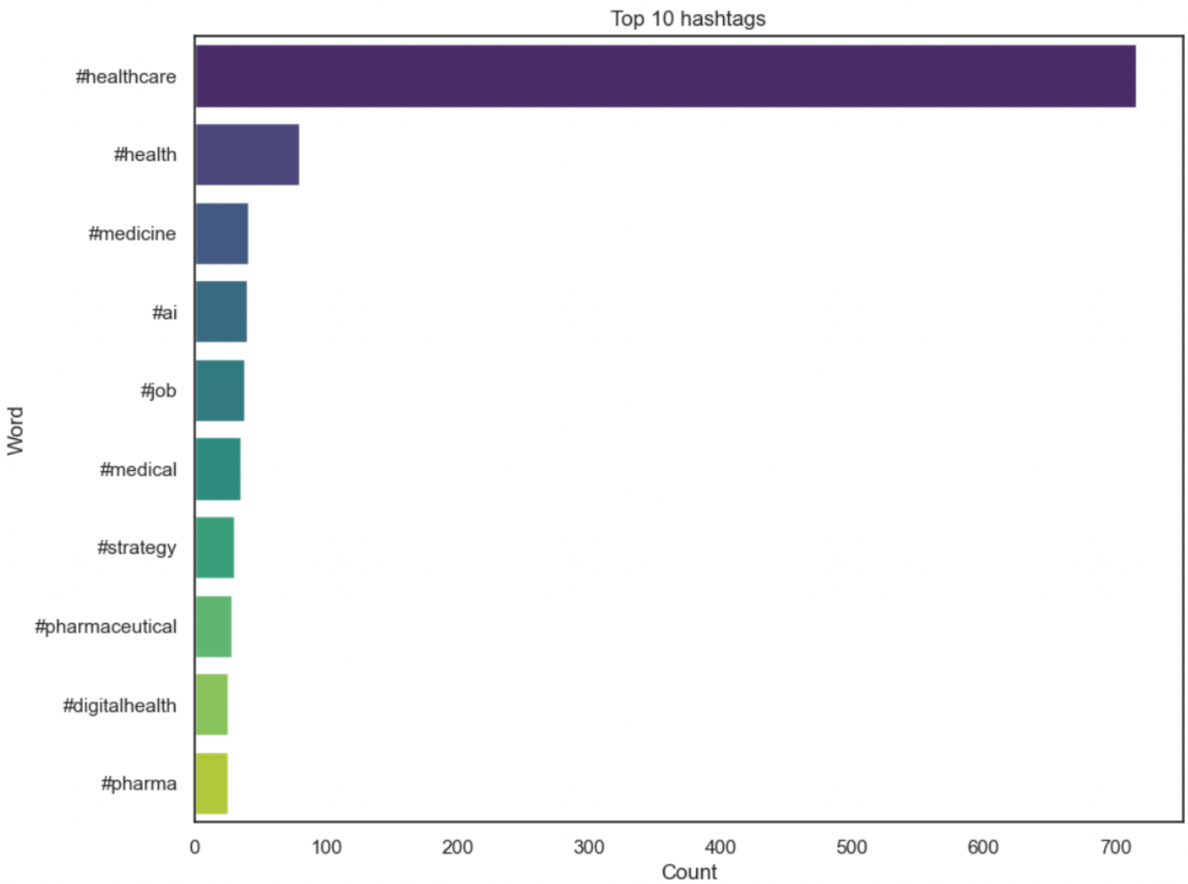

Figure 15: *Bar chart showing the top 10 hashtags.*

Interpretation:

- The most prevalent hashtag is ‘#healthcare’, followed by ‘#health’ and ‘#medicine’.

- The bar chart effectively displays the frequency of the top 10 hashtags used in X (Twitter) offering a concise comparison of their counts. Understanding hashtag usage is crucial for enhancing post engagement and highlighting popular topics.

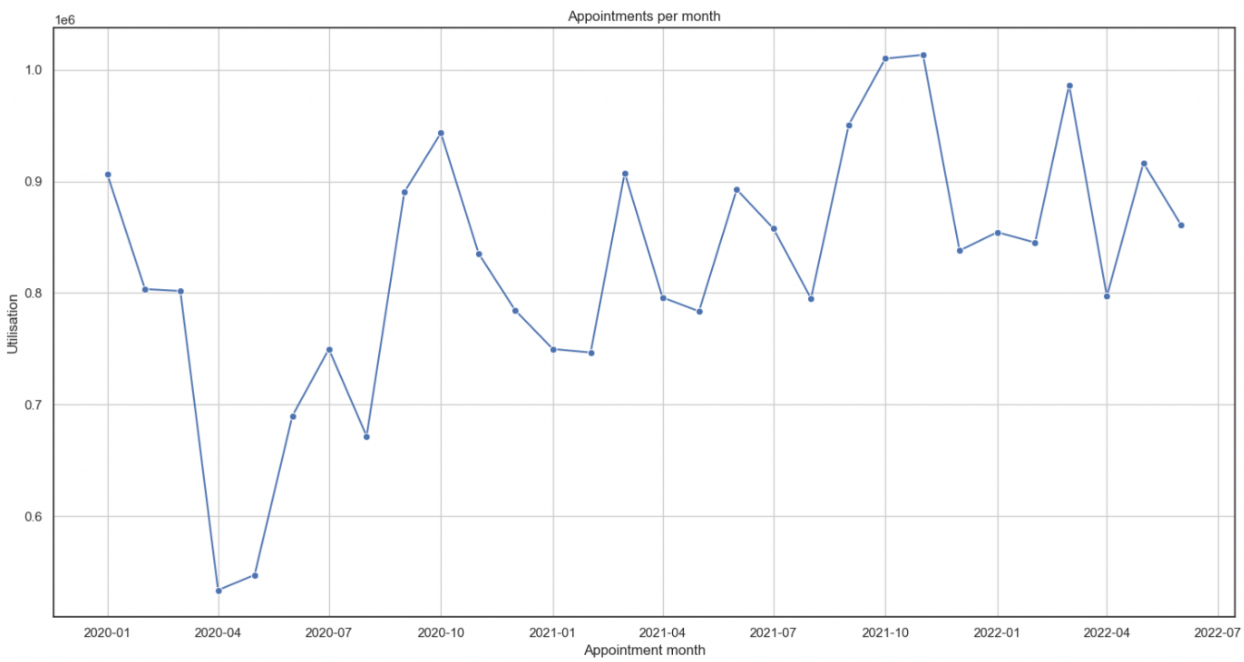

Figure 16: *Line chart showing the average appointments per day.*

Interpretation:

- The utilisation figures range from approximately 533,596 to 1,013,502, indicating the variation in appointment demand across different months.
- It appears that the NHS's planning guideline of 1,200,000 appointments per day aligns with or exceeds the observed utilisation in certain months, suggesting that the healthcare system is operating at or near its planned capacity.

<u>**Recommendations**</u>

Recommendations:

1. Standardised Data Entry practices: implement national standards for data entry across practice systems to enhance uniformity and reduce bias in data quality. Investment in a unified system can streamline data collection, for future exploration.
2. Training Programs: provide training on data entry protocols to healthcare staff, promoting accurate recording of appointment information and improving overall data accuracy.
3. Social Media Engagement: utilise social media platforms, such as (X) , to circulate information about the impact of missed appointments. This could look like creating awareness campaigns using engaging hashtags to discourage non-attendance.
4. Promotion of alternative appointment modes: encourage patients to book video/online appointments through social media promotion catering to those who are unable to attend physically, which may reduce the amount of missed appointments.
5. Budget allocation: Focus budget allocation on maximising existing infrastructure and resources within the NHS, ensuring more efficient healthcare delivery.
6. Observed appointment utilisation figures range from approximately 533,596 to 1,013,502. You may want to consider adjusting resources and staffing based on monthly variations to optimise efficiency to ensure that the NHS operates within or close to its planned capacity of 1,200, 000 appointments per day.

These recommendations aim to address data quality issues, enhance patient engagement, increase the system's responsiveness to varying demand levels and optimising resource allocation toward current infrastructure to help improve healthcare outcomes.

References

NHS England (2019). *NHS England» Missed GP appointments costing NHS millions*. [online] England.nhs.uk. Available at: https://www.england.nhs.uk/2019/01/missed-gp-appointments-costing-nhs-millions/.