

NHS: ANALYSIS OF APPOINTMENT DATA

AUGUST 2024



PREPARED BY

Henry Pescod

PREPARED FOR

NHS – LSE

Diagnostic Analysis of NHS Appointment Data using Python

Course 2 Assignment submission Henry Pescod - 12th August 2024

Table of Contents

1) Background/context of the business scenario	3
2) Analytical approach:	3
Section 1: Importing the necessary libraries and datasets provided and sensecheck	3
Section 2: Explore the data sets	4
3) Visualisation and insights:	5
Section 3. What is the actual utilisation of resources?	5
No of appointments by month:	5
Seasonal variation in Number of appointments by month:	6
Number of appointments by National Category:	10
Number of appointments by appointment mode:	11
Number of appointments by healthcare professional type:	12
Time between booking and appointments:	13
No of appointments by service setting:	14
Section 4. Is there adequate staff and capacity in the networks?	16
Average daily number of appointments per month:	16
Section 5. The problem of missed appointments	17
Count of missed appointments per month:	17
Number of missed appointments per month by appointment mode:	18
Number of missed appointments per month by health care professional type:	19
Number of Missed Appointments per Month by average waiting time:	20
Section 6. Identifying trends in X (FKA Twitter) engagement relating to healthcare in the UK	21
The most frequently used hashtags relating to healthcare in the UK:	21
4) Conclusions and recommendations:	22
Conclusions	22
Recommendations for further exploration	23
5) Appendix:	24
Background/context of the business scenario	24
Fishbone Diagram of Root Cause Analysis	24
Problem areas to be analysed in this report	25
Risks, assumptions and limitations	26
Analytical approach	27
Section 1: Importing the necessary libraries and datasets provided and sensecheck	27
Section 2: Exploring the data	27
Section 3: What is the actual utilisation of resources?	28
Section 4. Is there adequate staff and capacity in the networks?	28

Section 5. The problem of missed appointments	28
Section 6 - Identifying trends in X (FKA Twitter) engagement relating to healthcare in the UK	28
Visualisation and insights	28
Python functions used to do the visualisations	28
Patterns and predictions	29
Suggestions for further analysis and recommendations not included in the main report	29

1) Background/context of the business scenario

The NHS is the United Kingdom's Healthcare provider, managing millions of appointments each month. Through a data-driven analysis of appointments and resource utilisation the NHS aims to understand why patients missed GP appointments. The NHS incurs significant, potentially avoidable, costs when patients miss general practitioner (GP) appointments, and understanding the reasons better will support its future budget allocation. To attempt to tackle this problem, the NHS wants to understand its current resource utilisation, whether there is adequate staff and resources in the networks, and what this means for their X (twitter) engagement strategy.

A fishbone diagram was helpful to break the problem of why patients miss appointments [\[SEE APPENDIX\]](#) and determine which elements could be investigated in this analysis using the data sets provided.

2) Analytical approach:

Section 1: Importing the necessary libraries and datasets provided and sensecheck

Historical appointment data was analysed using Python and libraries like Pandas, NumPy, Matplotlib, and Seaborn to create DataFrames, perform exploratory analysis, and generate visualisations. The 3 data sets containing appointment data were provided as semi-wrangle files therefore required little in the way of cleaning. After importing the datasets into Python, they were sense checked using a variety of methods in Python [\[SEE APPENDIX\]](#). Using pivot tables in the dataset CSV files was also helpful to check the Python calculations were correct.

After an initiation exploration of the datasets provided using python I was able to extract these summary details.

Dataset	Time Period	No of Records	Total No of Appointments recorded	Key Data included

Actual Duration [AD]	1st December 2021 and 30th June 2022	137,793	167,980,692	Appointment duration Appointment location
National Category [NC]	1st August 2021 and 30th June 2022	817,394	296,046,770	Service setting Context type National category
Appointments Regional [AR]	1st January 2020 and 1st June 2022	596,821	742,804,525	Appointment status Hcp type Appointment mode Appointment waiting times

Section 2: Explore the data sets

After datasets were sense checked, the data sets were explored further using a list of questions and a variety of methods in Python [[SEE APPENDIX](#)].

The available data (highlighted in the key data column) was only going to be able to analyse some of [the possible factors linked to missed appointments](#).

Therefore I broke down the analysis into 4 parts:

What is the actual utilisation of resources?

Is there adequate staff and capacity in the networks?

The problem of missed appointments

Identifying trends in X (FKA Twitter) engagement relating to healthcare in the UK

Under each question a list of sub questions were identified to try and analyse the question further [[SEE APPENDIX](#)]. A set of limitations, assumptions of analysis in answering the problem were also thought through [[SEE APPENDIX](#)].

3) Visualisation and insights:

Section 3. What is the actual utilisation of resources?

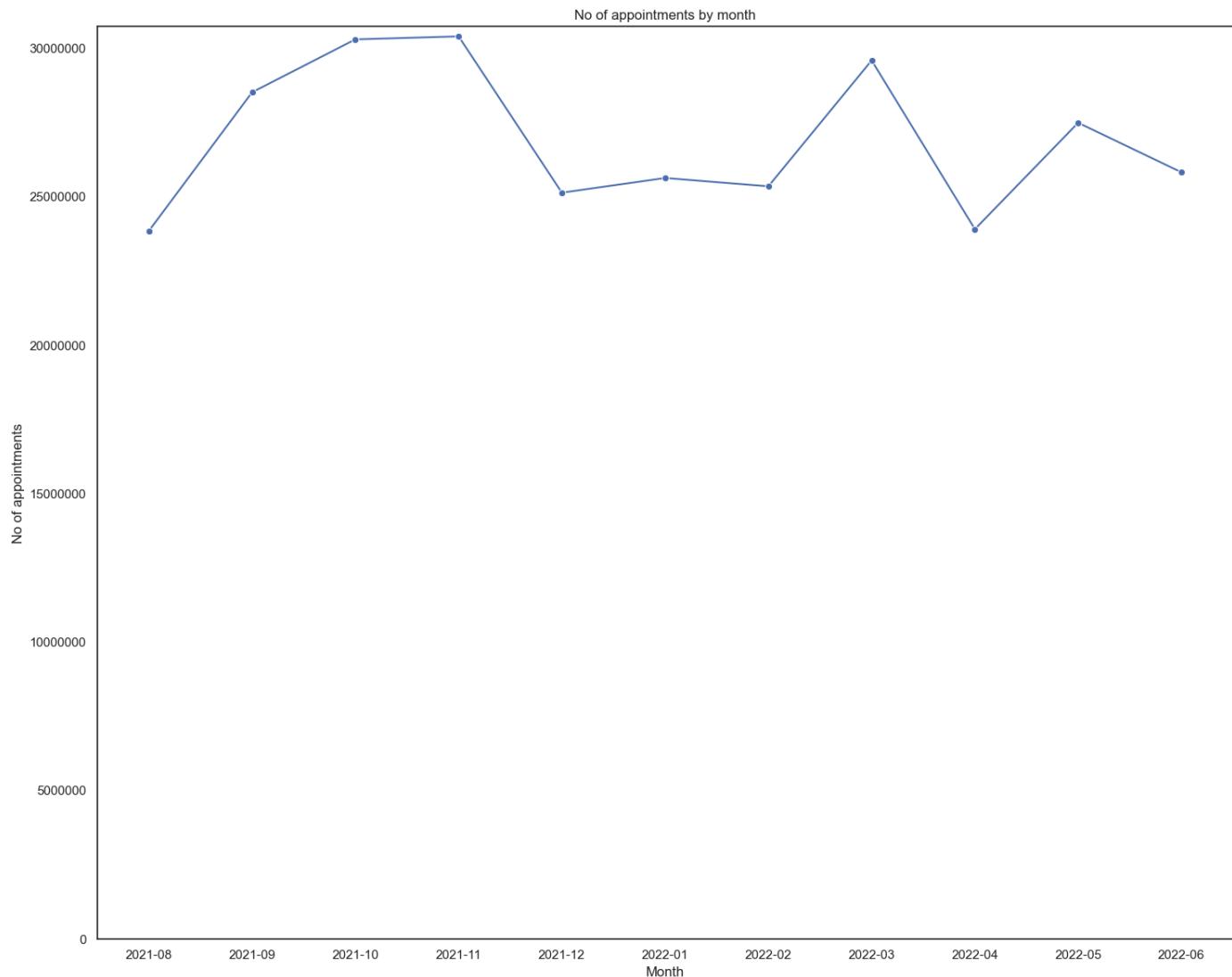
Section 3 and 4 of the analysis focuses on understanding the utilisation of resources and capacity in the networks. Based on the [problem analysis of looking for factors that affect missed appointments](#), the results could provide insight into:

- Booking flexibility: Could there be limited options for rescheduling or cancelling appointments due to capacity issues?
- Data Management: Is there poor data management leading to lost or miscommunicated appointment information?

A summary python methodology to create the visualisations can be found in [the appendix](#).

No of appointments by month:

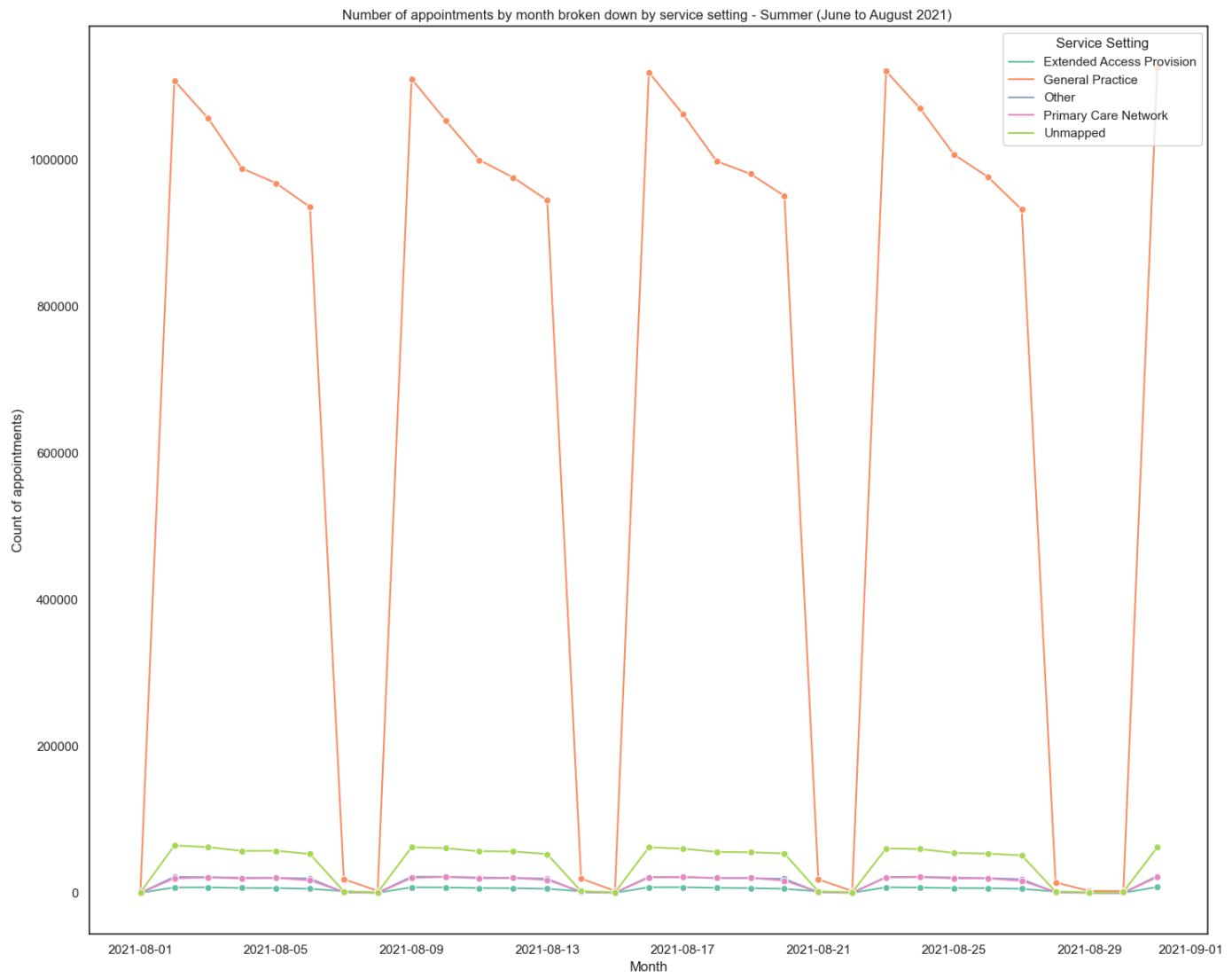
- There is significant seasonal variation in the number of appointments by month, for example November 2021 at 30,405,070, April with 2022 had 23,913,060 appointments.



Seasonal variation in Number of appointments by month:

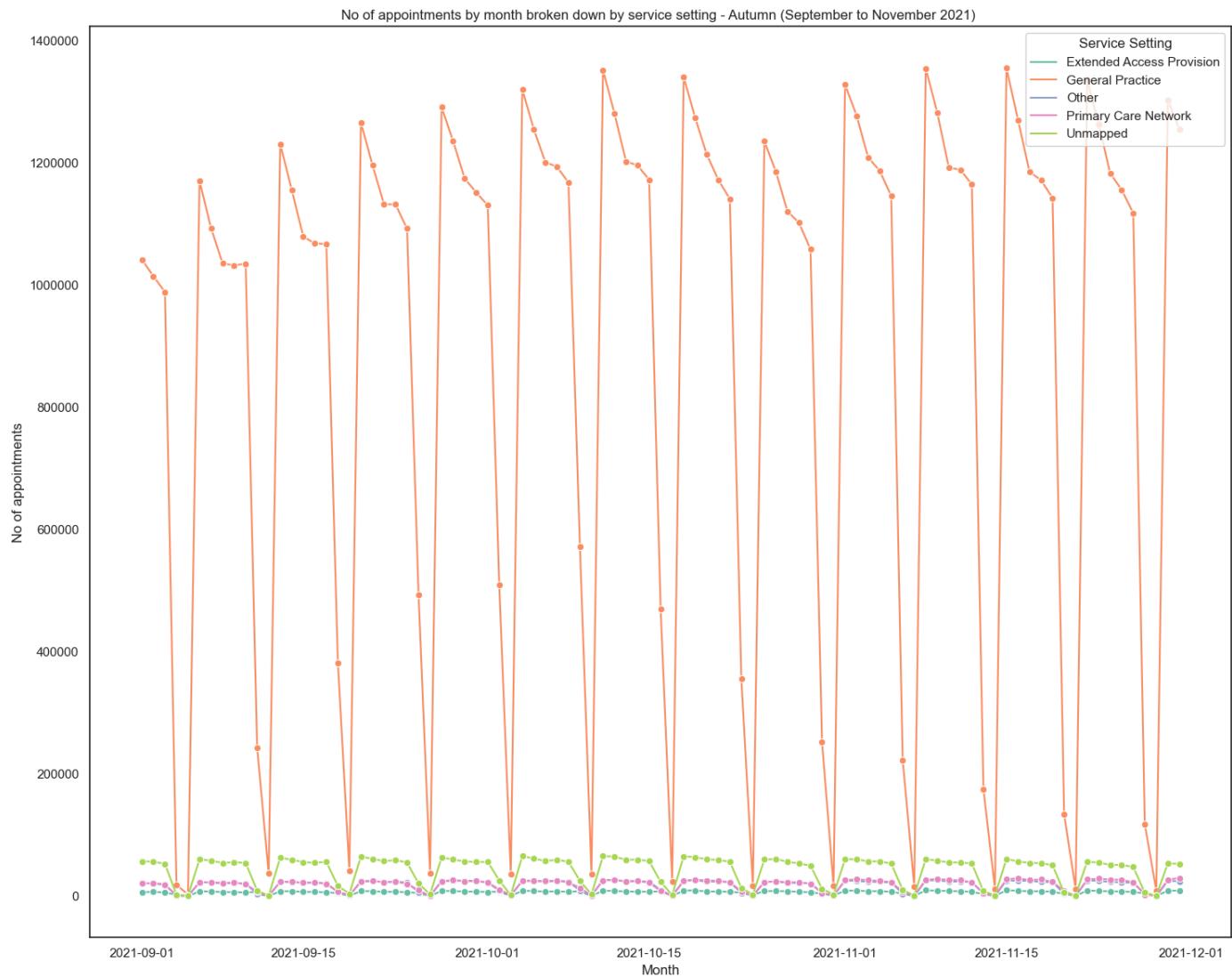
- The analysis then looked at the seasonal variations in number of appointments more closely to see what patterns of utilisation emerged.
- Summer 2011

Pescod_Henry_DA201_Assignment_Report.pdf



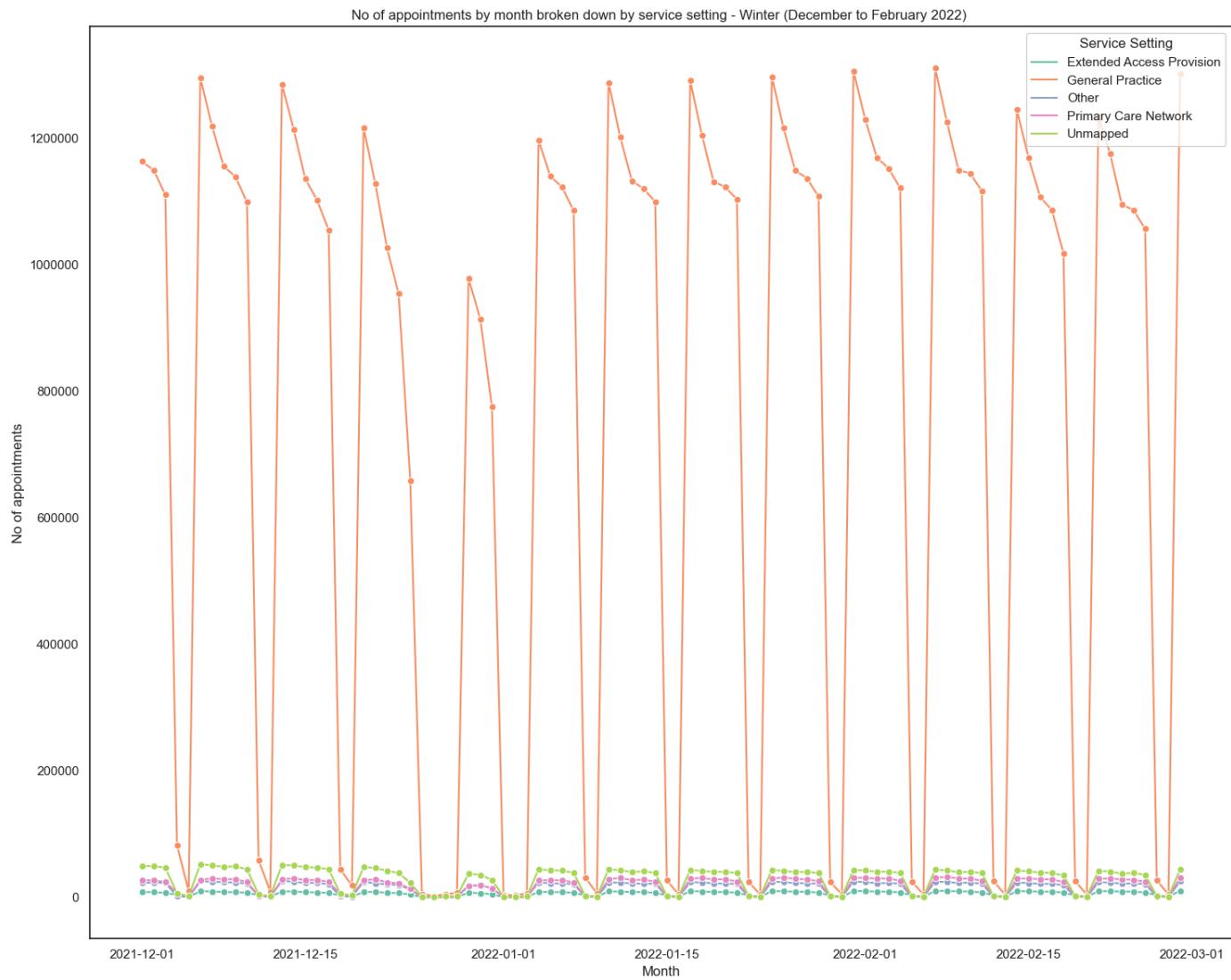
- Autumn 2021

Pescod_Henry_DA201_Assignment_Report.pdf

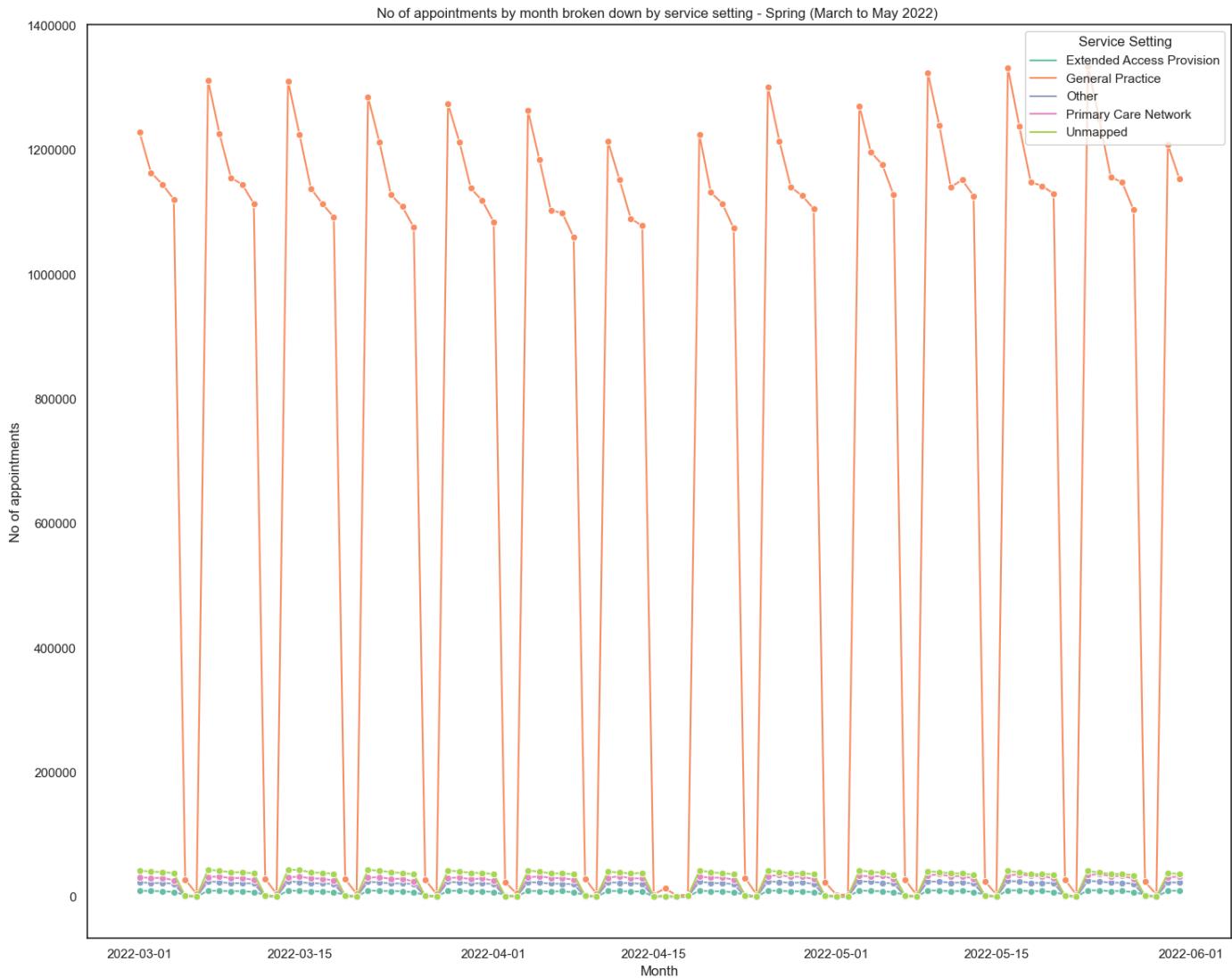


- Winter 2021 / 2022

Pescod_Henry_DA201_Assignment_Report.pdf



- Spring 2022

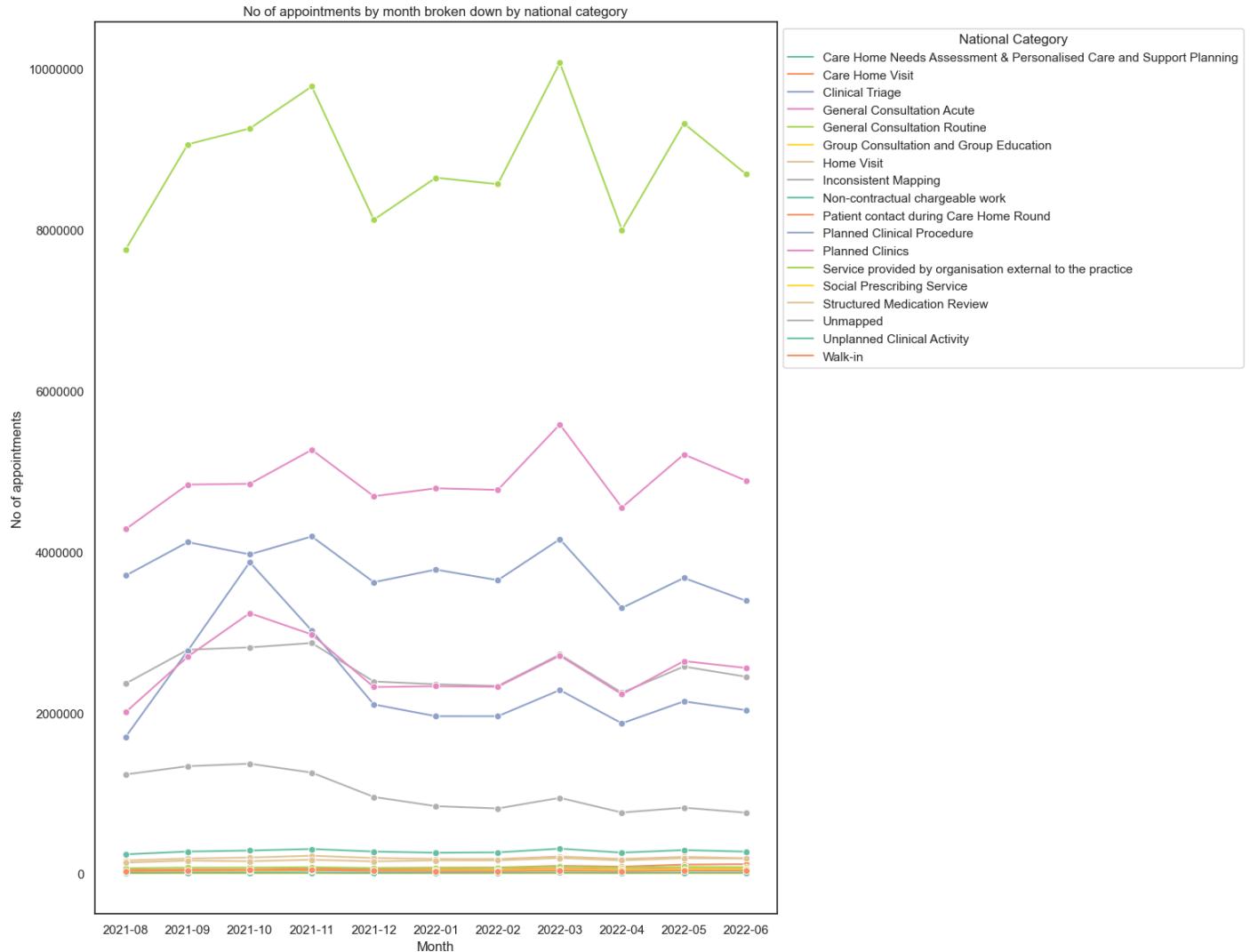


Some interesting insights

- In all seasons the number of appointments peak earlier in the week on a monday, and reduce over the week.
- No of appointments are at their lowest levels in the summer, and during holiday periods (Half terms, Christmas and Easter)
- Autumn experiences sees a steady increase in number of appointments leading up to December which should be factored in to resources planning

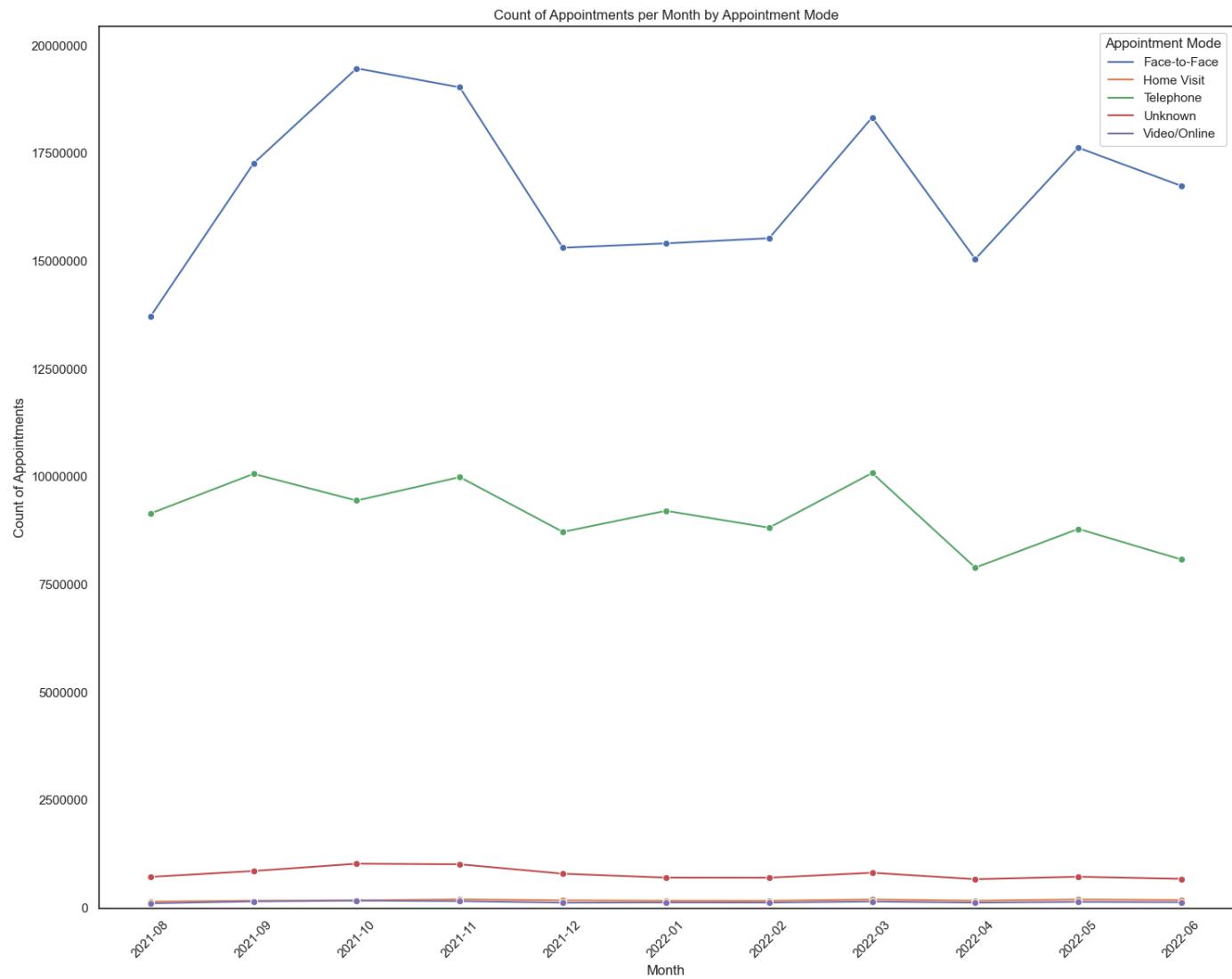
Number of appointments by National Category:

- A third of all appointments in the NC dataset are classed as General Consultation Routine. 13% of appointments are incorrectly mapped, which suggests some further investigation could be made by NHS into data handling practices.



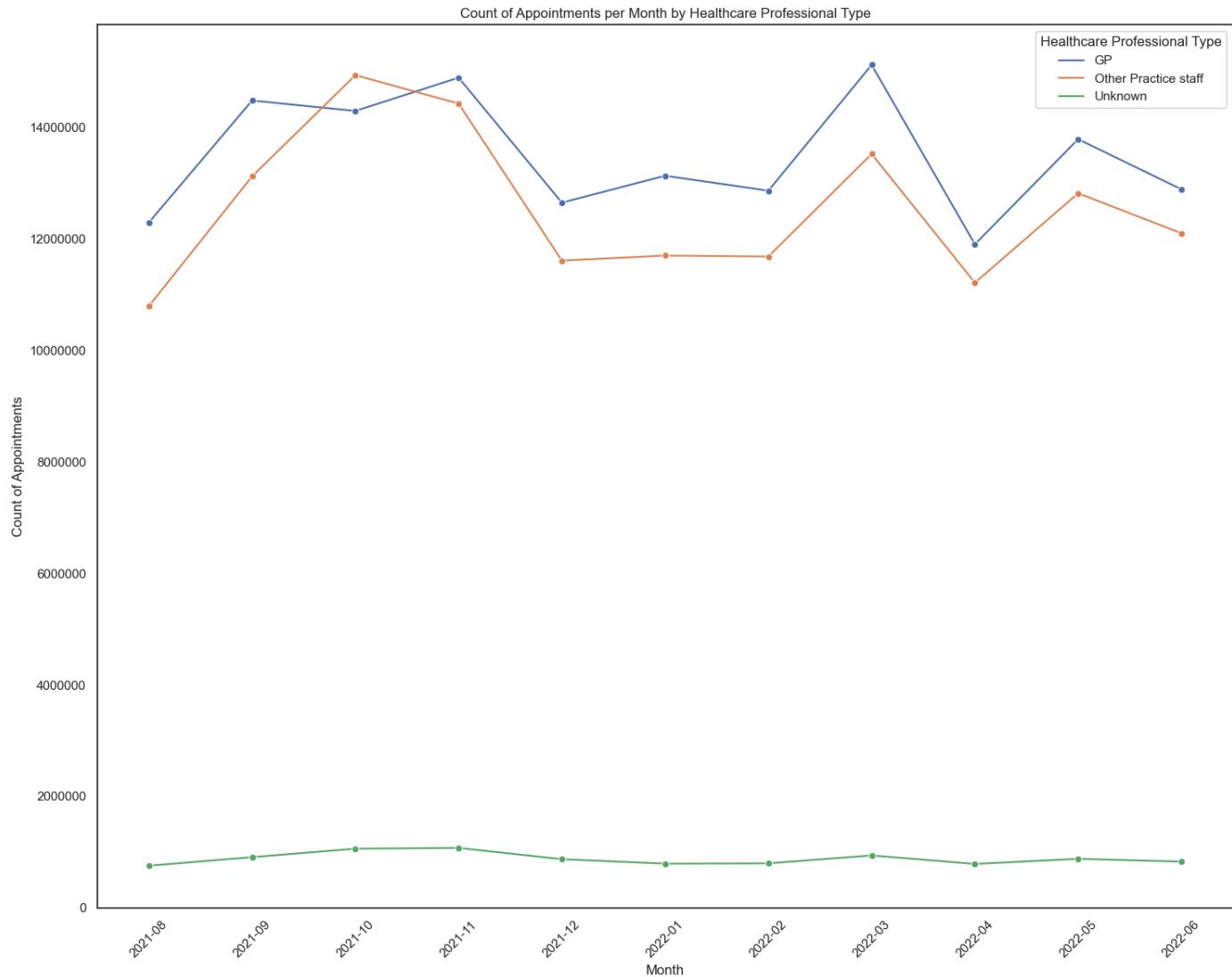
Number of appointments by appointment mode:

- Face to face appointments make up the bulk of appointments, following by telephone appointments
- Very few appointment are conducted by Video / Online



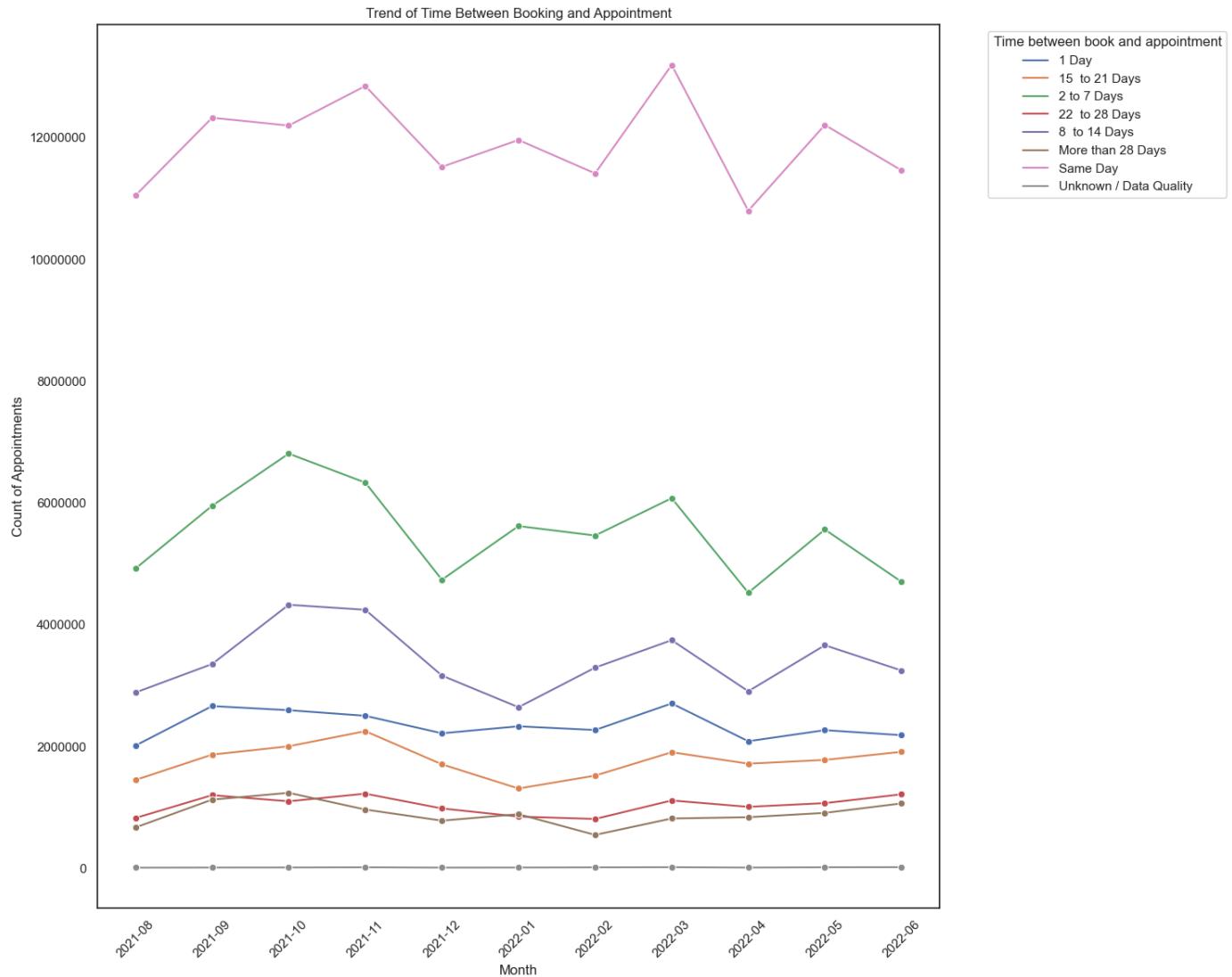
Number of appointments by healthcare professional type:

- The number of appointments are split fairly evenly between GP and other practice staff



Time between booking and appointments:

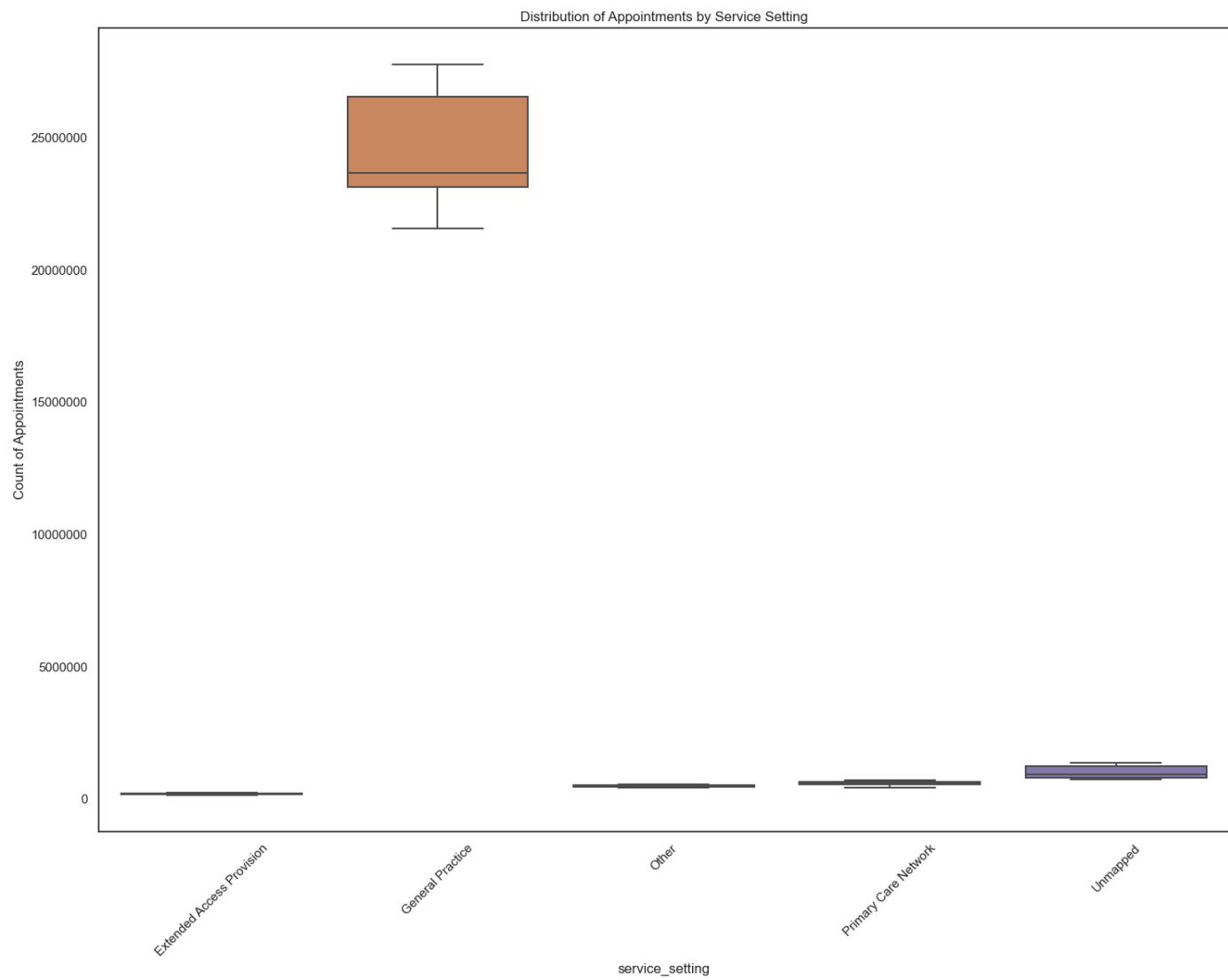
- Same day appointments account for the most number of appointments
- 2 to 7 days and 8 to 14 days are more frequent waiting times than next day appointments.

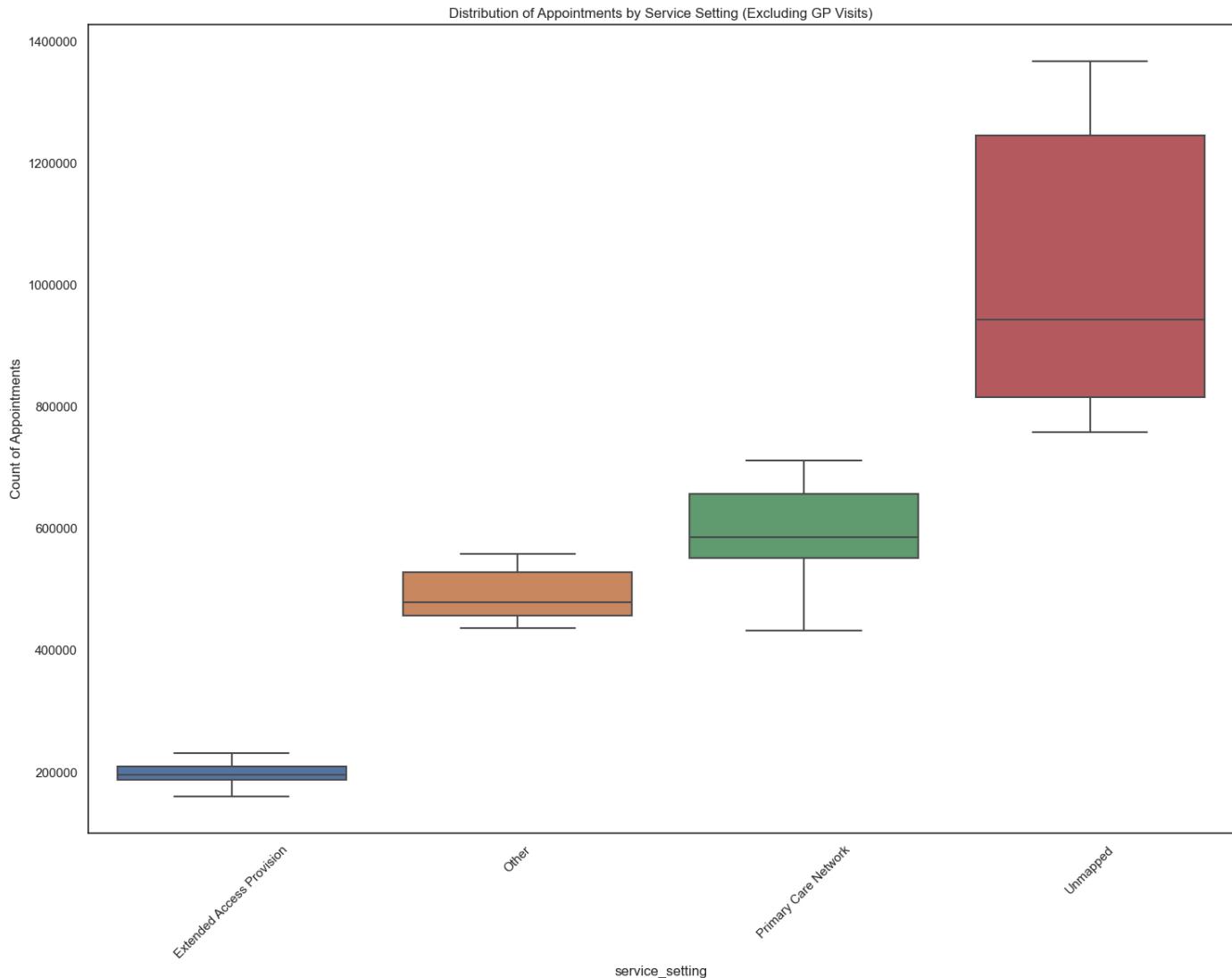


No of appointments by service setting:

- General Practice service setting accounts the majority share of total appointments with an Interquartile Range between 23 - 27m appointments per month
- Large number of unmapped service settings is the data which could suggest investing in improving data management practices

Pescod_Henry_DA201_Assignment_Report.pdf

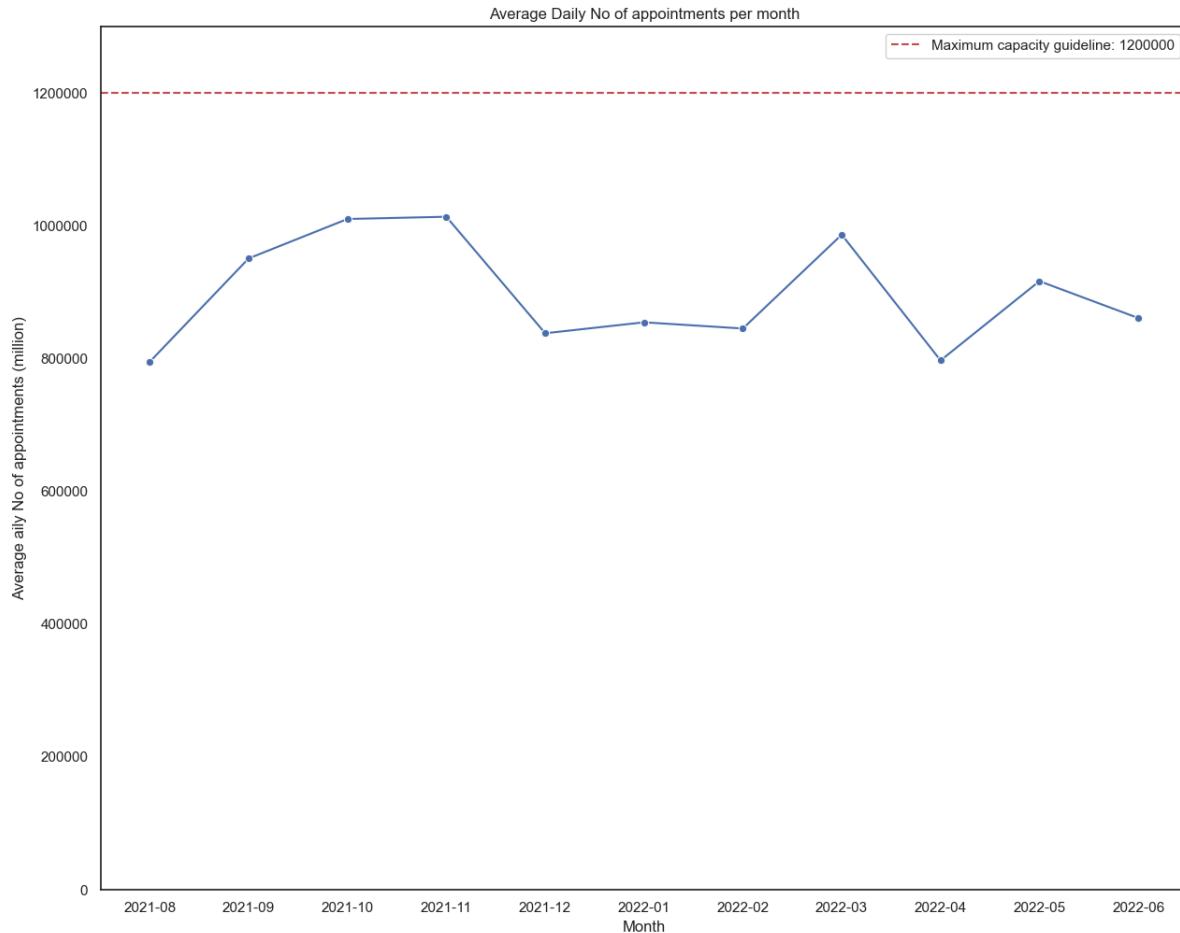




Section 4. Is there adequate staff and capacity in the networks?

Average daily number of appointments per month:

- November 2021 had the highest average daily appointments which did not go above the NHS daily appointment capacity guideline of 1,200,000, suggesting there is currently sufficient capacity in the networks.



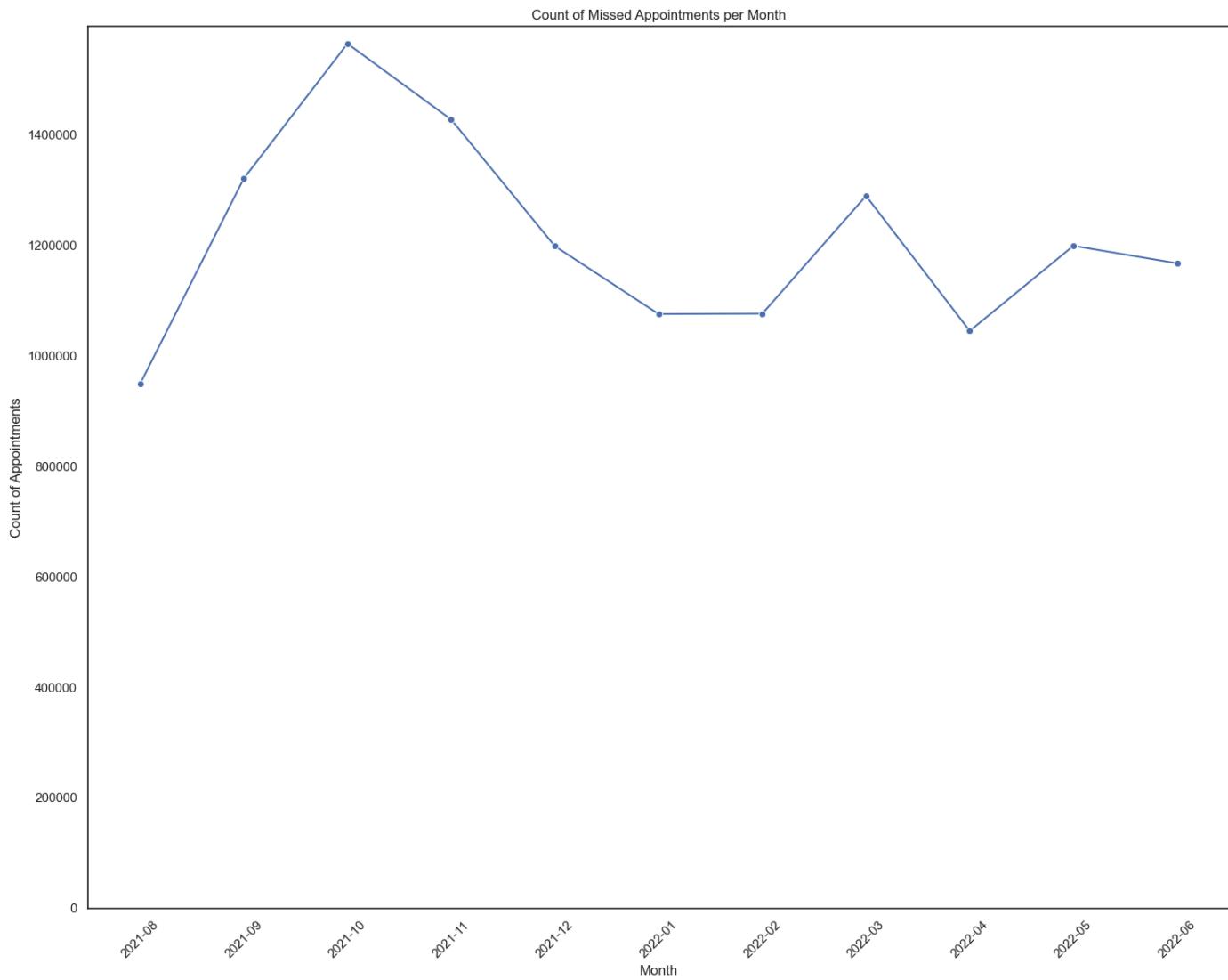
Section 5. The problem of missed appointments

This section of analysis focuses the following [possible problem areas](#):

- Accessibility: Do appointment types impact missed appointments due to location difficulty?
- Training: Are certain healthcare professionals missing more appointments due to inadequate training?
- Resource Allocation: Is there a link between missed appointments and the resources for patient communication and reminders, especially regarding waiting times?

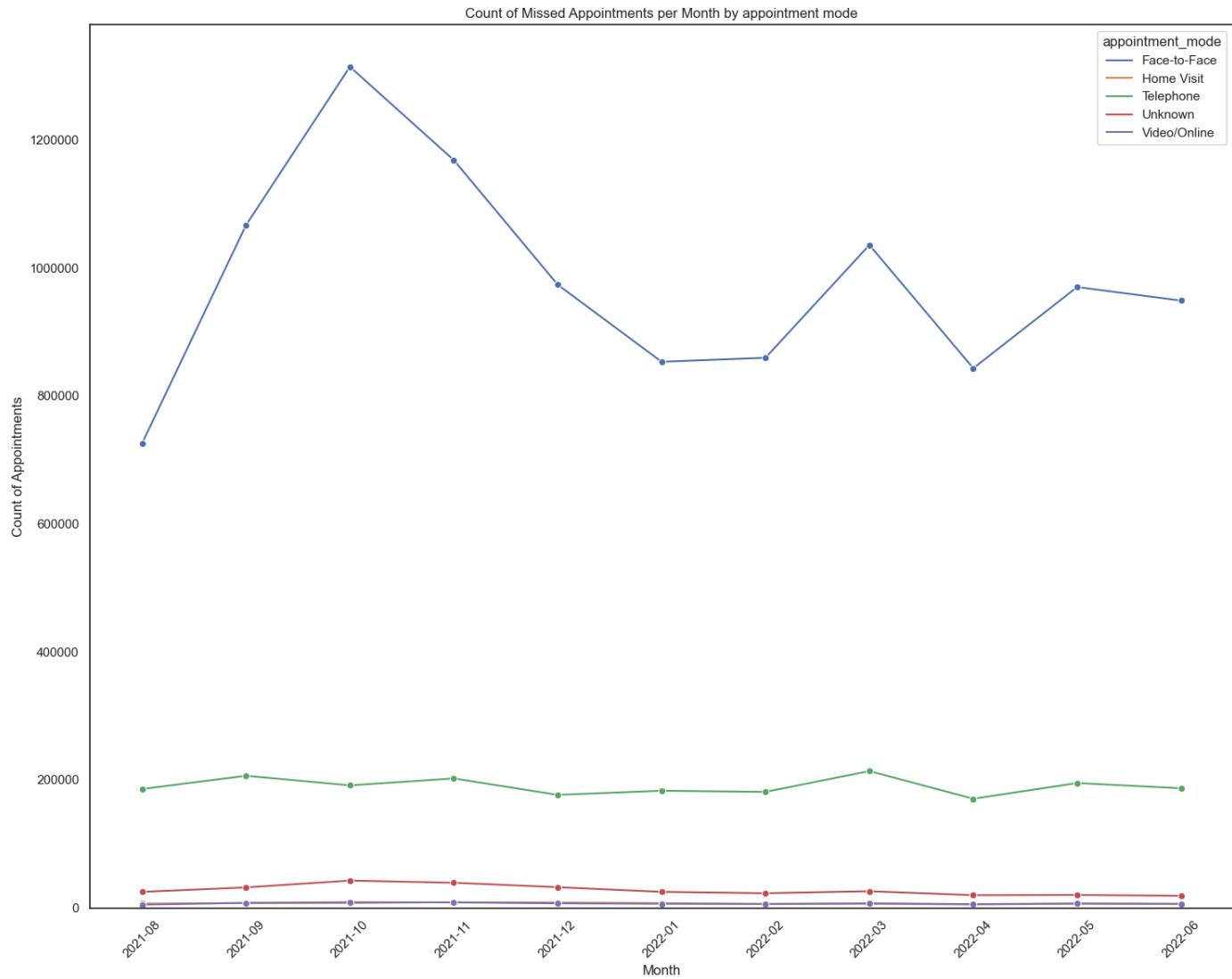
Count of missed appointments per month:

- There is a higher number of missed appointments in months with more overall appointments.
- More appointments are missed in Autumn / Winter months than in spring / summer months.



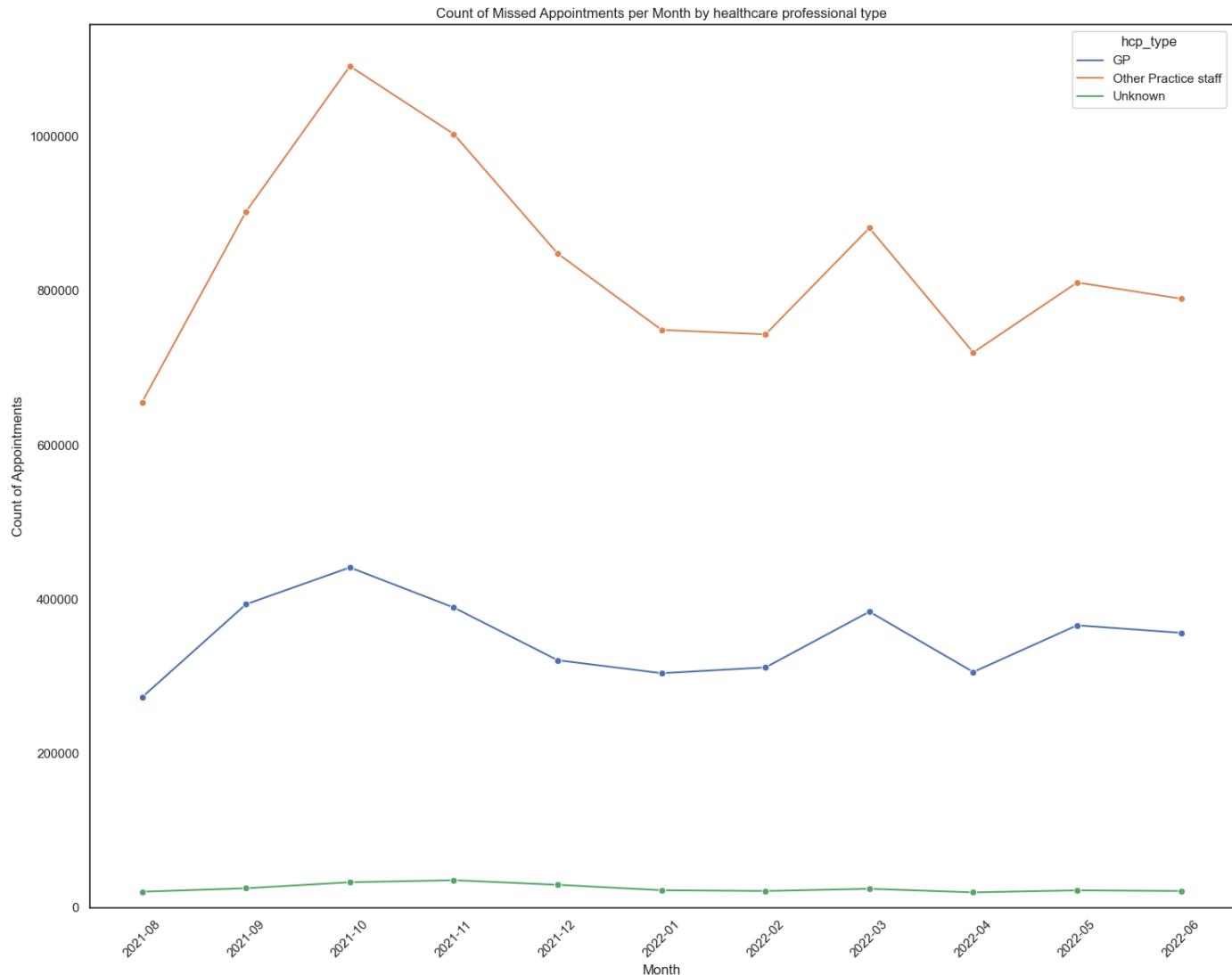
Number of missed appointments per month by appointment mode:

- Face to face appointments are the most common type of missed appointment.
- Comparing this chart to the total count of appointments by appointment mode, its visible number of appointments missed by telephone is proportionally lower compared to face to face appointments



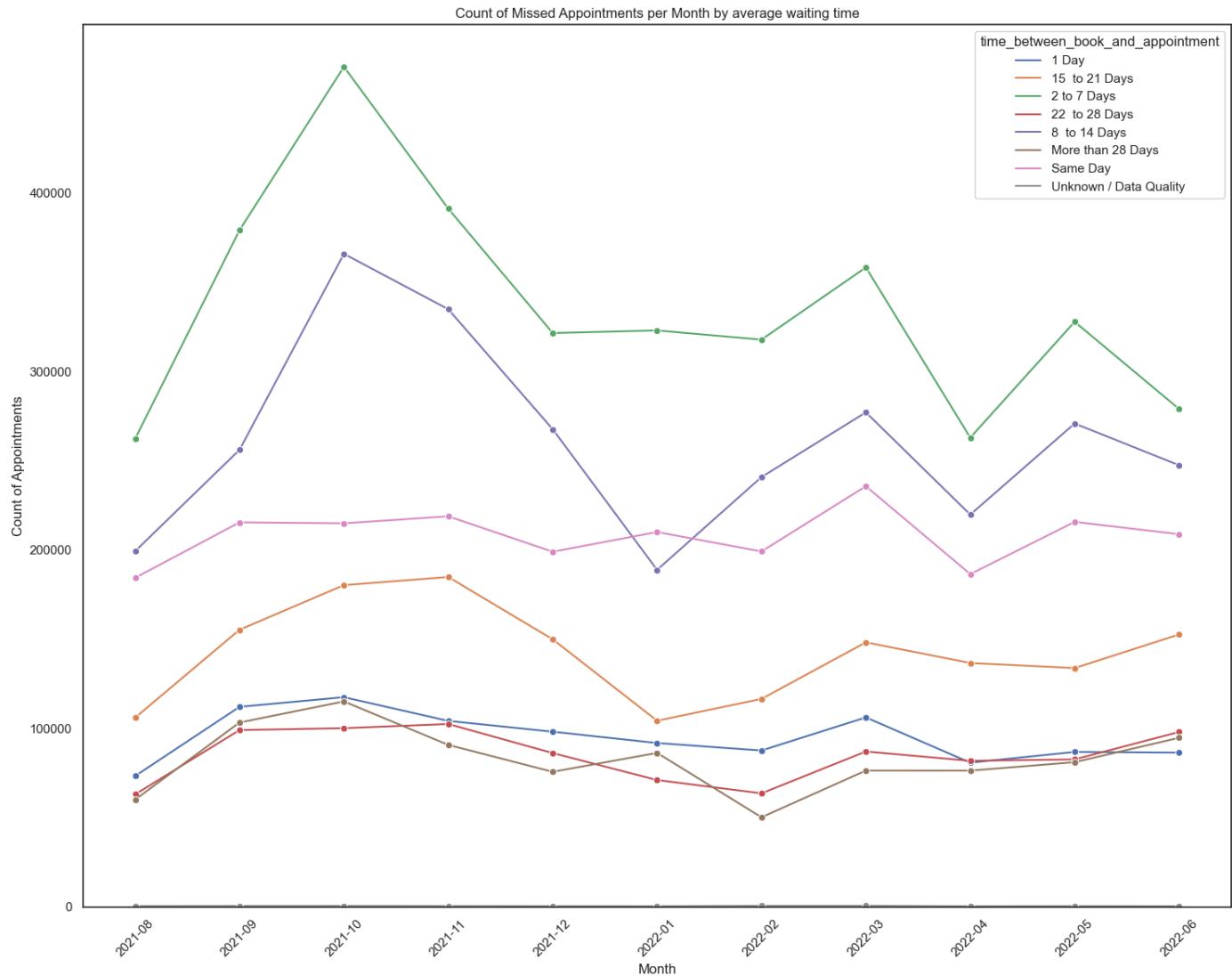
Number of missed appointments per month by health care professional type:

- The number of missed appointments is significantly higher for other practice staff than GPs, which is interesting because they have a similar total number of appointments as seen in [section 3](#).



Number of Missed Appointments per Month by average waiting time:

- 2 to 7 day and 8-14 day waiting time experience the most missed appointments. Suggests longer waiting periods are more likely to be missed -
- Same day appointments are missed proportionately less compared to number of appointments they account for



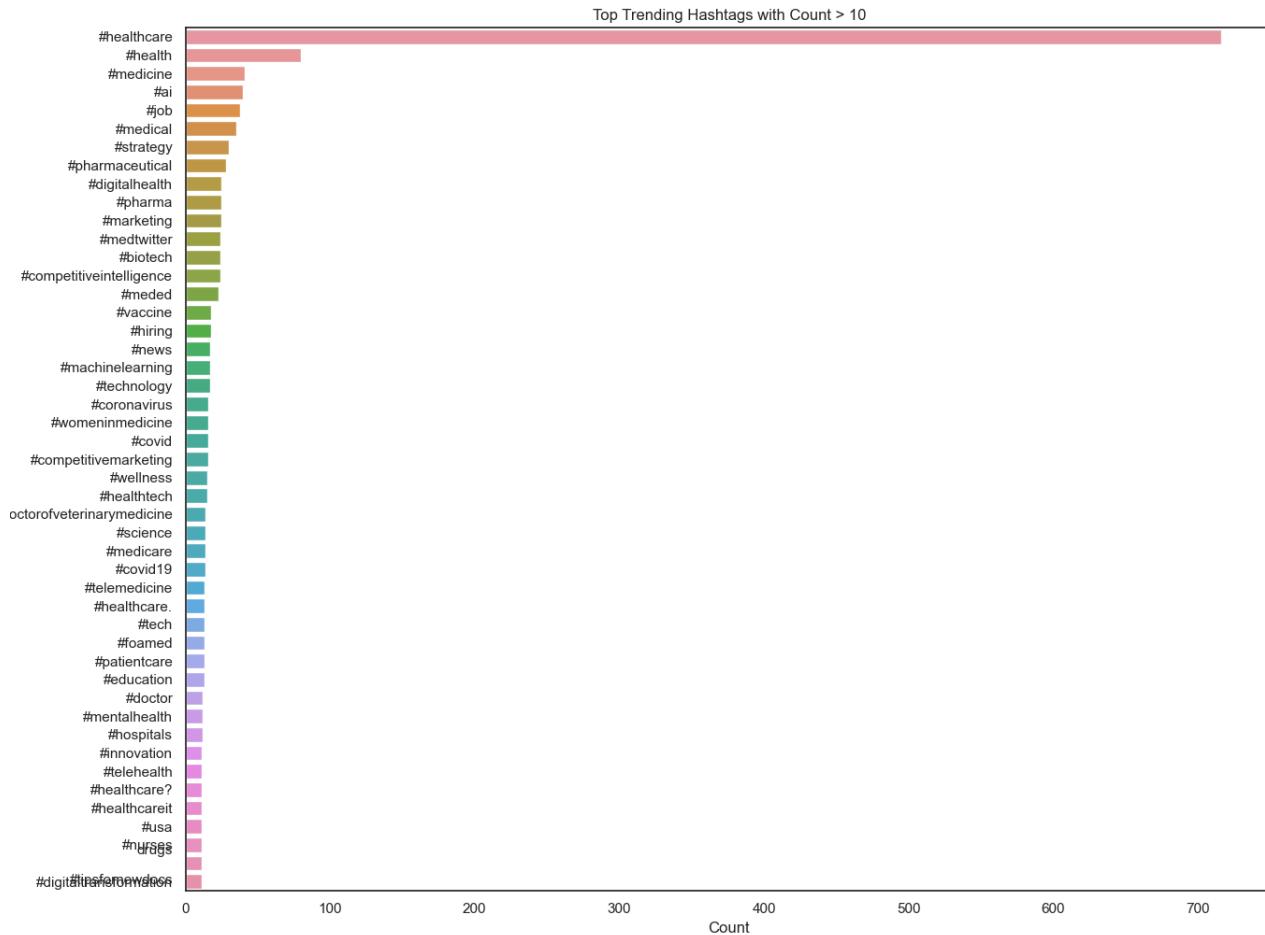
Section 6. Identifying trends in X (FKA Twitter) engagement relating to healthcare in the UK

One of the possible factors identified affecting missed appointments is awareness, i.e. a lack of understanding about the importance of keeping appointments. The analysis in this section looks at tweets analysis and how this could be used to raise awareness about the importance of keeping appointments.

The most frequently used hashtags relating to healthcare in the UK:

- The most frequently used hashtag is #healthcare with 716 occurrences, followed by #health (80), #medicine (41), and #ai (40).

- The NHS should consider using these popular hashtags to increase the reach and engagement of their tweets.



4) Conclusions and recommendations:

Conclusions

1. There is no capacity issue in the networks suggesting the problem of missed appointments is not related to a capacity problem.
2. Same-day appointments, despite being the most frequent, were missed less often than those with longer waiting periods.
3. Telephone appointments have a much lower % rate of missed appointments compared to face to face appointments

4. A significantly higher proportion of appointments are missed when held by other practice staff rather than GPs when they have a similar total number of appointments.
5. The most frequently used hashtag is #healthcare with 716 occurrences, followed by #health (80), #medicine (41), and #ai (40).

Recommendations for further exploration

6. Can the NHS conduct more Telephone or Video / Online appointments which have a lower rate of being missed than face to face appointments
7. The NHS should explore how appointment reminders are issued to patients with longer appointment waiting periods who miss proportionately more appointments.
8. The NHS should explore why more appointments are missed by other practice staff.
9. The NHS should consider using the most popular hashtags to increase the reach and engagement of their tweets. An example could participate in conversations around #ai if NHS are using any AI-based solutions to reduce missed appointments

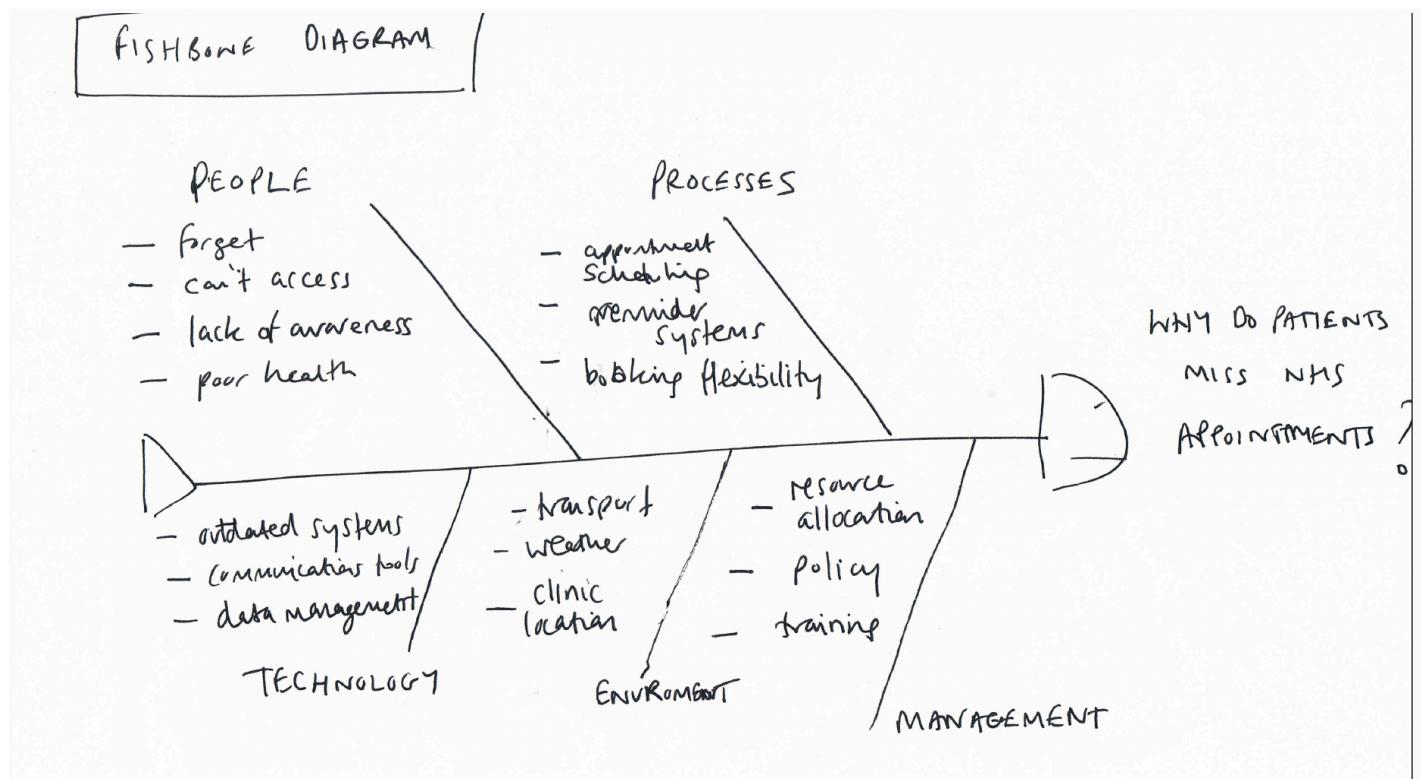
Further recommendations can be found in [the appendix](#).

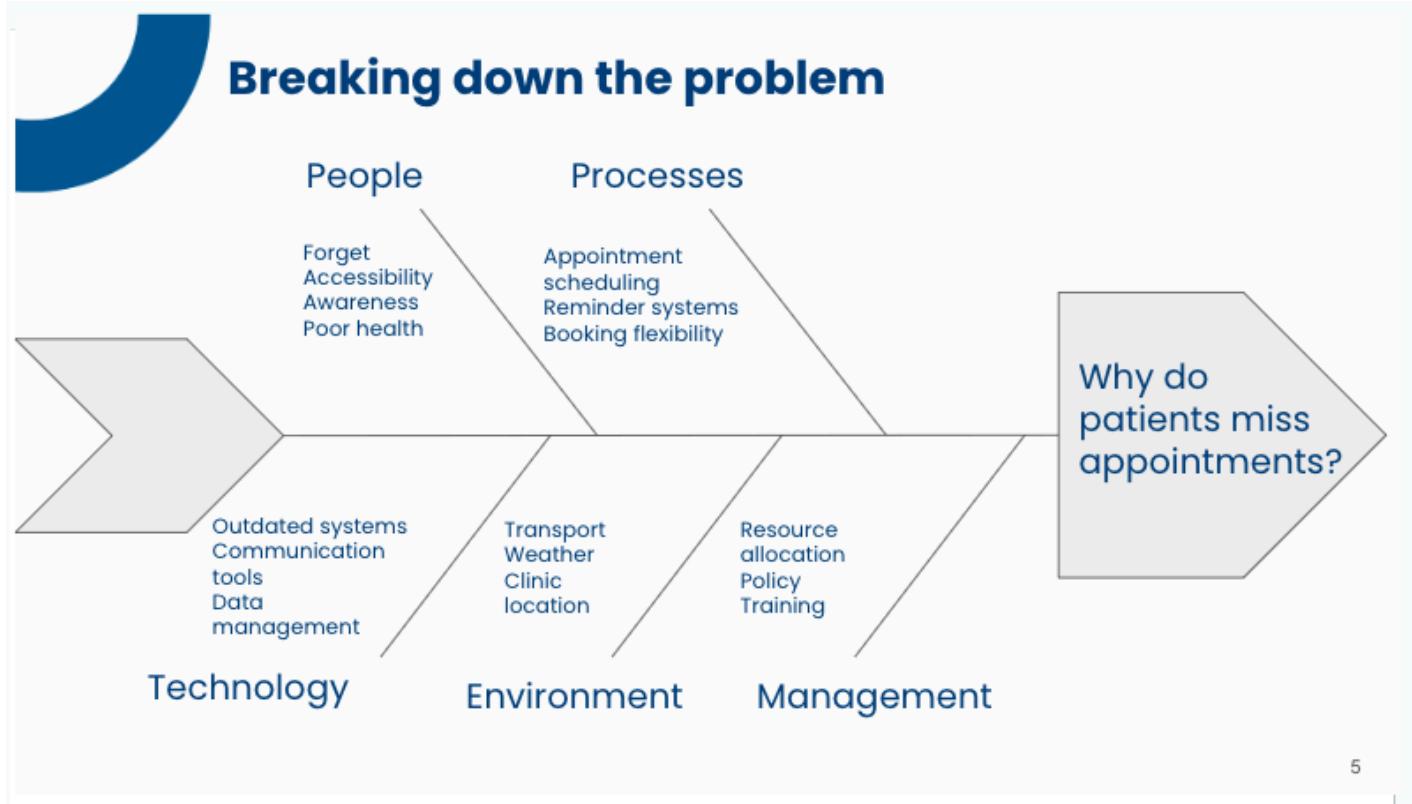
Word Count = 1,072

5) Appendix:

Background/context of the business scenario

Fishbone Diagram of Root Cause Analysis





People

- **Forgetfulness:** Patients may forget their appointments.
- **Accessibility:** Difficulty in reaching the appointment location.
- **Awareness:** Lack of understanding about the importance of keeping appointments.
- **Health Issues:** Sudden illness or worsening of condition on the appointment day.

Processes

- **Appointment Scheduling:** Complicated or inconvenient scheduling processes.
- **Reminder Systems:** Ineffective or non-existent reminder systems.
- **Booking Flexibility:** Limited options for rescheduling or cancelling appointments.

Technology

- **Outdated Systems:** Inefficient or outdated appointment scheduling and reminder systems.
- **Communication Tools:** Lack of integration with modern communication tools like SMS, email, or app notifications.
- **Data Management:** Poor data management leading to lost or miscommunicated appointment information.

Environment

- **Transport:** Poor public transportation options or long travel times.

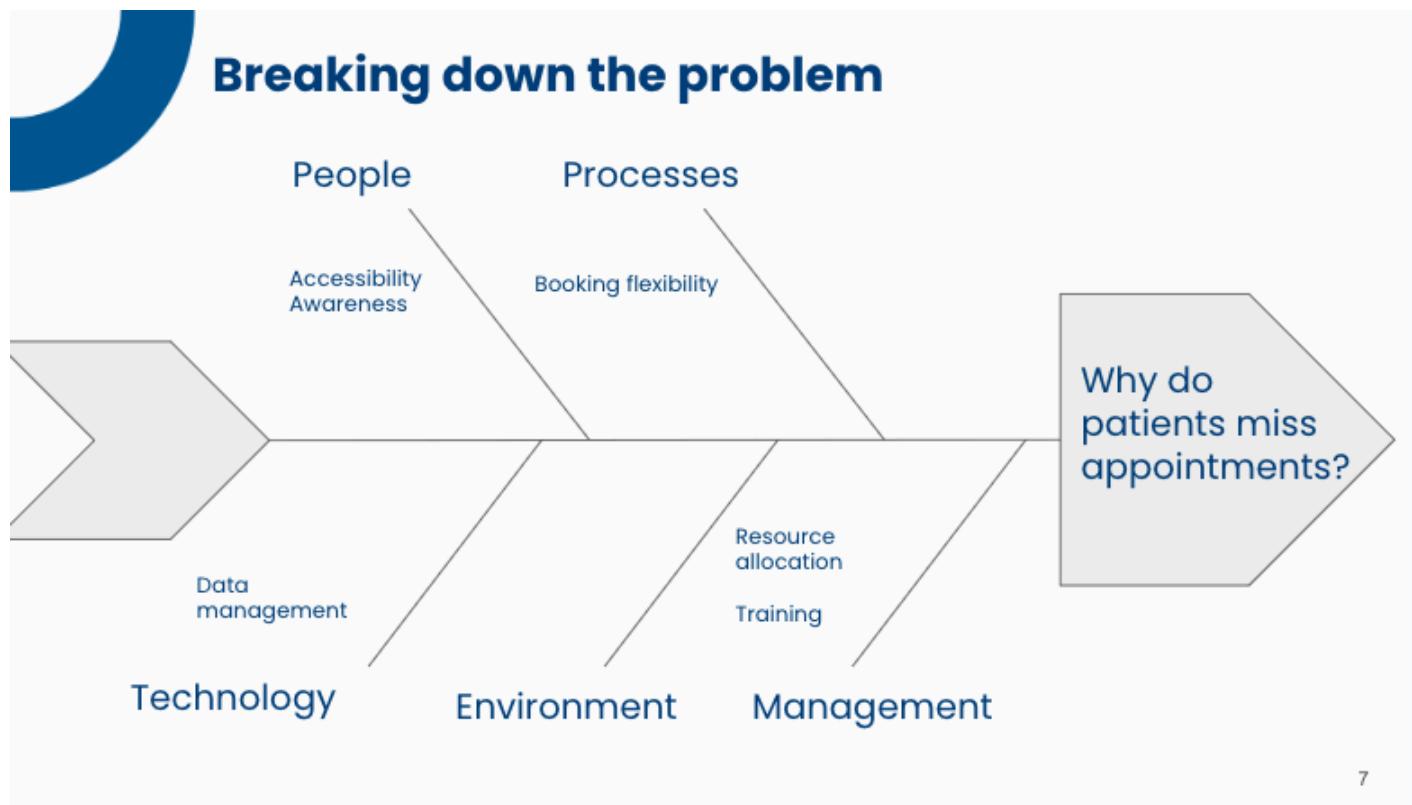
- **Weather Conditions:** Bad weather deterring patients from attending.
- **Facility Location:** Clinics or hospitals located far from residential areas.

Management

- **Resource Allocation:** Insufficient resources allocated to patient communication and reminders.
- **Policy:** Lack of policies promoting flexible appointment management.
- **Training:** Insufficient training for staff on managing patient appointments and follow-ups.

Problem areas to be analysed in this report

The sections in yellow - highlight which area of the problem which will be explored and analysed in this report using the datasets provided.



- **Accessibility:** Difficulty in reaching the appointment location. Analysis look at how missed appointments are affected by type of appointments (appointment_mode)

Awareness: Lack of understanding about the importance of keeping appointments. Analysis looks at tweets analysis and how this could be used to raise awareness.

Booking Flexibility: Limited options for rescheduling or cancelling appointments. Part of this problem could be investigated when looking at the utilisation of resources and capacity in the networks.

Data Management: Poor data management leading to lost or miscommunicated appointment information. Analysis looks at how much data is unmapped in the datasets.

Resource Allocation: Insufficient resources allocated to patient communication and reminders. The analysis looks at utilisation of resources and capacity in the networks.

Training: Insufficient training for staff on managing patient appointments and follow-ups. Analys could look at which healthcare professionals miss the most appointments

Problem area:	Accessibility	Booking Flexibility	Resource Allocation	Data Management	Training	Awareness
Possible problem	Difficulty in reaching the appointment location.	Limited options for rescheduling or cancelling appointments	Insufficient resources allocated to patient communication and reminders.	Poor data management leading to lost or miscommunicated appointment information	Insufficient training for staff on managing patient appointments and follow-ups.	Lack of understanding about the importance of keeping appointments.
Analysis question	How is the number of missed appointments affected by type of appointment?	Could there be limited options for rescheduling or cancelling appointments due to capacity issues?	How do the number of missed appointments differ by average waiting time?	How much data is unmapped in the datasets?	Which healthcare professionals miss the most appointments?	How could X(twitter) be used to raise awareness about missed appointment more effectively?

Risks, assumptions and limitations

Limitations of the analysis: The fishbone breakdown of the problem does also provide a useful framing to show the number of different factors affecting the problem of missed appointments, and why further investigation into the other areas of the problem not covered in this report will be required to gather the full picture.

Datasets: Some key decisions were focussing on the problem of missed appointments and the factors behind this, the analysis on missed appointments, therefore the analysis focuses more on the NC, AR and tweets datasets.

Limitations of the time period:

- Focussed on time period 1st August 2021 and 30th June 2022 from the NC dataset to be able to compare by NC and AR over the same time period. This did limit the seasonal breakdown of appointments in summer 2021 which only covered August 2021.

- Also any insights and recommendations taken from these data sets should bear in mind the short time period of the data. The results should be compared with other time periods to get a better sense of the mean, median and variation of the results.
- The time periods are over COVID period which may adversely affect the data (i.e a higher number not able to attend face to face appointments).

Joining datasets: The decision was made not to merge or join the datasets due to differing time periods covered and different categories recorded in each.

Analytical approach

Section 1: Importing the necessary libraries and datasets provided and sensecheck

Python functions used to sense check the data sets

- info() to see column names, data types
- isna() to check for missing values
 - A missing value check did identify some missing values in the tweets.csv file but these were deemed not to affect data quality.
- describe() to view the descriptive statistics for each data set

Some interesting initial observations from exploring the 3 datasets on counts of appointments:

Location: Top 5 locations (4.7% of total 106 locations) account for 16% of the total number of appointments, with 4 of the top 5 locations based in the south east of England.

Appointment Status: 91% of appointments in the NC dataset are classed as attended. 4% of appointments in the nc dataset are classed as missed. Section 5 of this report will look at missed appointments in more detail.

Section 2: Exploring the data

List of questions to explore the data

Q1: How many locations are there in the data set?

Q2: What are the five locations with the highest number of appointments?

Q3: How many service settings, context types, national categories, and appointment statuses are there?

Q4: Between what dates were appointments scheduled?

Q5: Which service setting reported the most appointments in North West London from 1 January to 1 June 2022?

Q6: Which month had the highest number of appointments?

Python functions used to explore the data sets

- Value_counts to identify the number of unique locations
- Groupby & sum() to aggregate the count of appointments by various categories
- To_datetime to convert to date format for ease of analysis
- To_period to extract year and month from date columns when analysing figures by month
- User defined function to run multiple aggregation based on service setting, but filtered for different date
 - def aggregate_by_day_and_service(data):
 - return data.groupby(['appointment_date', 'service_setting'])['count_of_appointments']\
 - .sum().reset_index()

Section 3: What is the actual utilisation of resources?

Q7: How do the number of appointments per month vary for service settings, context types, and national categories.

Q8: How do the number of appointments vary by season

Section 4. Is there adequate staff and capacity in the networks?

Q9: utilisation capacity - Should the NHS start looking at increasing staff levels?

Q10: hcp type - How do the healthcare professional types differ over time?

Q11: appointment status - Are there significant changes in whether or not visits are attended?

Q12: appointment mode - Are there changes in terms of appointment mode and the busiest months?

Q13: time between booking and appointment - Are there any trends in time between booking and appointment?

Q14: How do the various service settings compare?

Section 5. The problem of missed appointments

Q15: How does the trend of missed appointments vary over time?

Q16: How does the trend of missed appointments vary by appointment mode?

Q17: How does the trend of missed appointments vary by average waiting time for appointments?

Section 6 - Identifying trends in X (FKA Twitter) engagement relating to healthcare in the UK

Q18: What do the tweet_retweet_count and tweet_favorite_count columns say about levels of engagement?

Q19: What can be learnt from the top performing tweets based on tweet_retweet_count and tweet_favorite_count?

Q20: What can be learnt from the most frequently used hashtags?

Visualisation and insights

Python functions used to do the visualisations

- Sns.lineplot () for creating line plots
- sns.boxplot() for creating boxplots
- plt.tight_layout() to make the visualisations fit the png files
- plt.xticks(rotation=45) to rotate x-axis labels for better readability
- plt.ticklabel_format(style='plain', axis='y') to rescale the tick labels on the y axis to prevent scientific notation.
- ax.set_ylim(bottom=0) to set the y-axis limit to start at 0

Patterns and predictions

Suggestions for further analysis and recommendations not included in the main report

Further analysis could look into how the time of year and weather conditions affects missed appointments, and whether this should change the number of appointment modes offered during these months

Further analysis could look at mapping population data to the appointment datasets to look for any interesting trends between population density and number of appointments.

Further analysis could be considered to look at missed appointments broken down by national category - this would require joining of the AR and NC datasets

Should facilitating more next day appointments be considered?

Further analysis could look into benchmarking analysis comparing these figures to countries with similar demographics and healthcare systems.

Pescod_Henry_DA201_Assignment_Report.pdf