NHS Report

Scenario Context

The National Health Service (NHS) is a publicly funded healthcare system in the United Kingdom. The government would like to know how their resources have been allocated and utilised over a certain period (January 2020 - June 2022) in order to gain a better understanding of why people miss appointments.

Exploratory analysis of the provided data will uncover and present findings with which the project leader can use to inform decision-making related to the project goal of reducing the number of missed appointments.

Questions for the Department of Health and Social Care

• Would it be possible to access data more directly related to NHS staff?

Further Questions of the Data

• Could more information be gleaned from investigating the causes of data being recorded as **Unmapped** or **Unknown**?

Analytical Approach

- I imported:
    - **pandas**
    - **numpy**
    - **datetime**
    - **matplotlib.pyplot**
        - **matplotlib.ticker**
    - **seaborn**
    - **warnings**

- Duplication was central to the nature of the patient-related datasets and so I decided not to remove duplicate entries.
    - This assumption may have led to less accurate analysis.

- I decided not to merge the data sets.
    - Although **sub_icb_code** could be used as a merge key, the codes in each DataFrame contain differing levels of detail with regard to time data and appointment counts for each record, so merging could have produced incorrect or misleading data for analysis.
    - Computer processing / memory capacities proved an issue. I would have had to merge small subsets, but I wanted to see a broader picture as this was exploratory research.
        - This decision may have limited potential analytical insights with regard to comparing variables from separate data sets.

- I determined the locations (n = 106) with the highest number of appointments.
    - Using **.groupby()** with **.agg()** is an efficient and effective way to filter, group, and aggregate data.
        1. **North West London** (n = 12142390)
        2. **North East London** (n = 9588891)
        3. **Kent and Medway** (n = 9286167

- Within each of these locations I determined the **service setting**, **context type**, and **national category** with the highest number of appointments so as to gain a broad understanding of where NHS resources were being allocated.

```
1   # Create a subset of the nc DataFrame.
2   nc_NWL = nc.groupby(['sub_icb_location_name', 'service_setting',
3                        'appointment_date'])\
4                       ['count_of_appointments'].agg('sum').reset_index()
5
6   # Filter subset for specific date range.
7   nc_NWL = nc_NWL[(nc_NWL['appointment_date'] > "2022-01-01")
8                       & (nc_NWL['appointment_date'] < "2022-06-01")]
9
10  # Filter subset for specific location.
11  nc_NWL = nc_NWL[nc_NWL['sub_icb_location_name']\
12          .str.contains("NHS North West London ICB - W2U3Z")]
13
14  # Sort the subset DataFrame to view which service setting
15  # reported the most appointments.
16  nc_NWL.sort_values(by = ['count_of_appointments'], ascending=False)
```

- I expanded the date selection and investigated the prevalence of each variable.

    - Most appointments occurred within:

        - Service Setting: **GP**

        - Context Type: **Care Related Encounter**

        - National Category: **General Consultation Routine**

• I determined which months had the highest number of appointments overall.

```
1   # Use groupby() to subset the ar DataFrame with
2   # appointment_month and count_of_appointments.
3                                           # Incorporate .agg() function
4                                           # to add the appointments.
5   nc_gpby_app = nc.groupby(['appointment_month'])\
6                       ['count_of_appointments'].agg('sum').reset_index()
7
8   # Sort the subset to see the highest number of appointments.
9   nc_gpby_app.sort_values(by = 'count_of_appointments', ascending=False)
```

- The month with the highest number of appointments was **November 2021**.
- The month with the highest number of records was **March 2022**.

Patterns, Trends & Insights

- I created visualisations by aggregating data on a national level for a broader view of resource allocation.
    - As the project progresses, visualisations could be produced for specific locations.

- As **General Practice** has already been shown to have significantly more appointments than the other service settings, I plotted it in isolation to see a clearer picture of how each changed over time.

## Appointments for General Practice Over Time
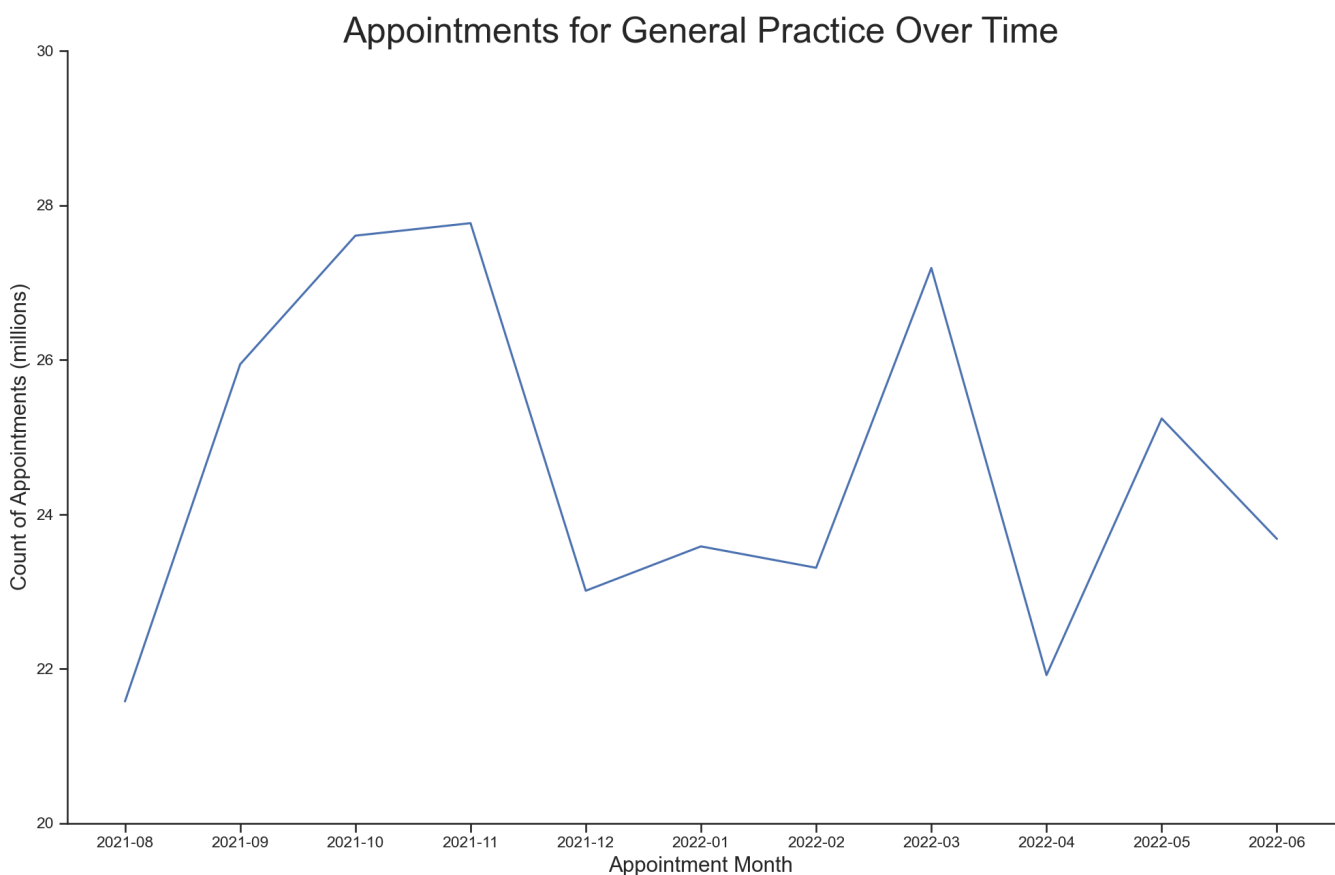
- This plot trend aligns with the previous findings that the run-up to Christmas 2021 and **March 2022** are the busiest periods.

Appointments for Service Settings (Excluding GP) Over Time
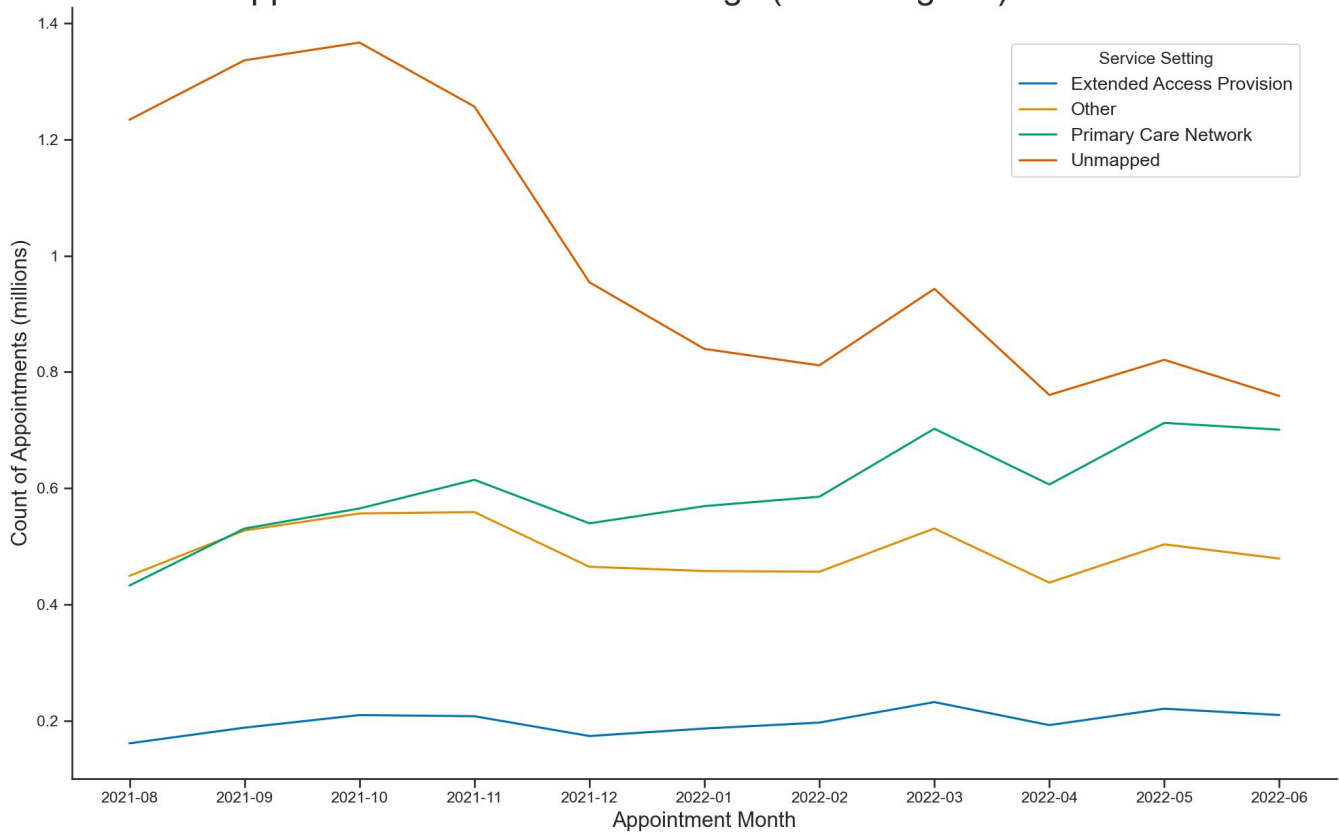
- The rise in number of appointments for the other service settings are, in relative terms, also significant increases.


The Spread of Appointments for Each Service Setting (Excluding GP)
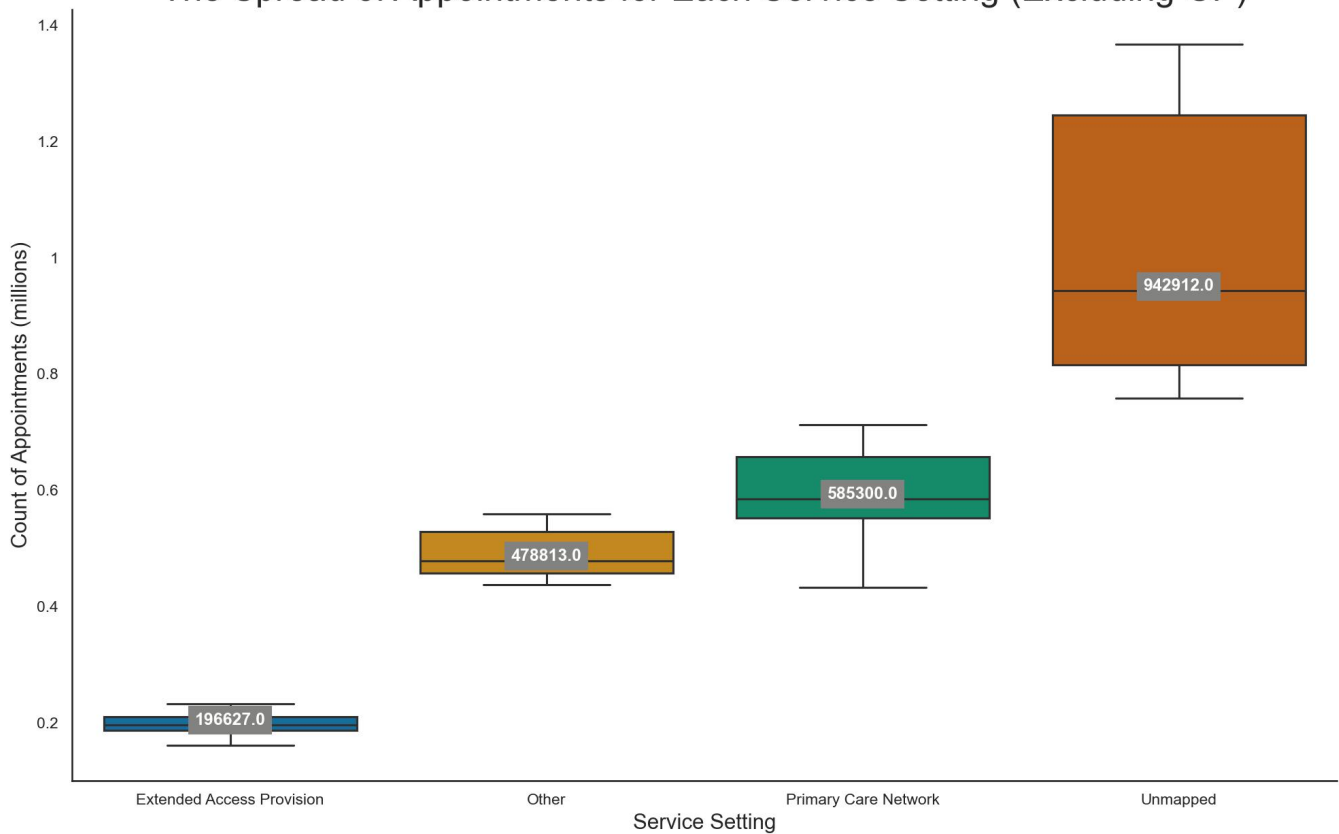
- The **Unmapped** category contains more records than the others.
  - It would be useful to understand why this is, as the true mapping of these data points may change the distribution significantly.
    - Could it be due to poor data collection?
      - Was data collection only poor during the months recorded in these datasets?
        - Analysing data from preceding and succeeding months could answer this question.

- Plotting all **National Categories** (n = 18) together is aesthetically *busy*, so it is beneficial from the perspective of looking to see where resources have been most utilised, to filter for **National Categories** that have more appointments.

Appointments for National Categories Over Time (Apps > 1.2m)
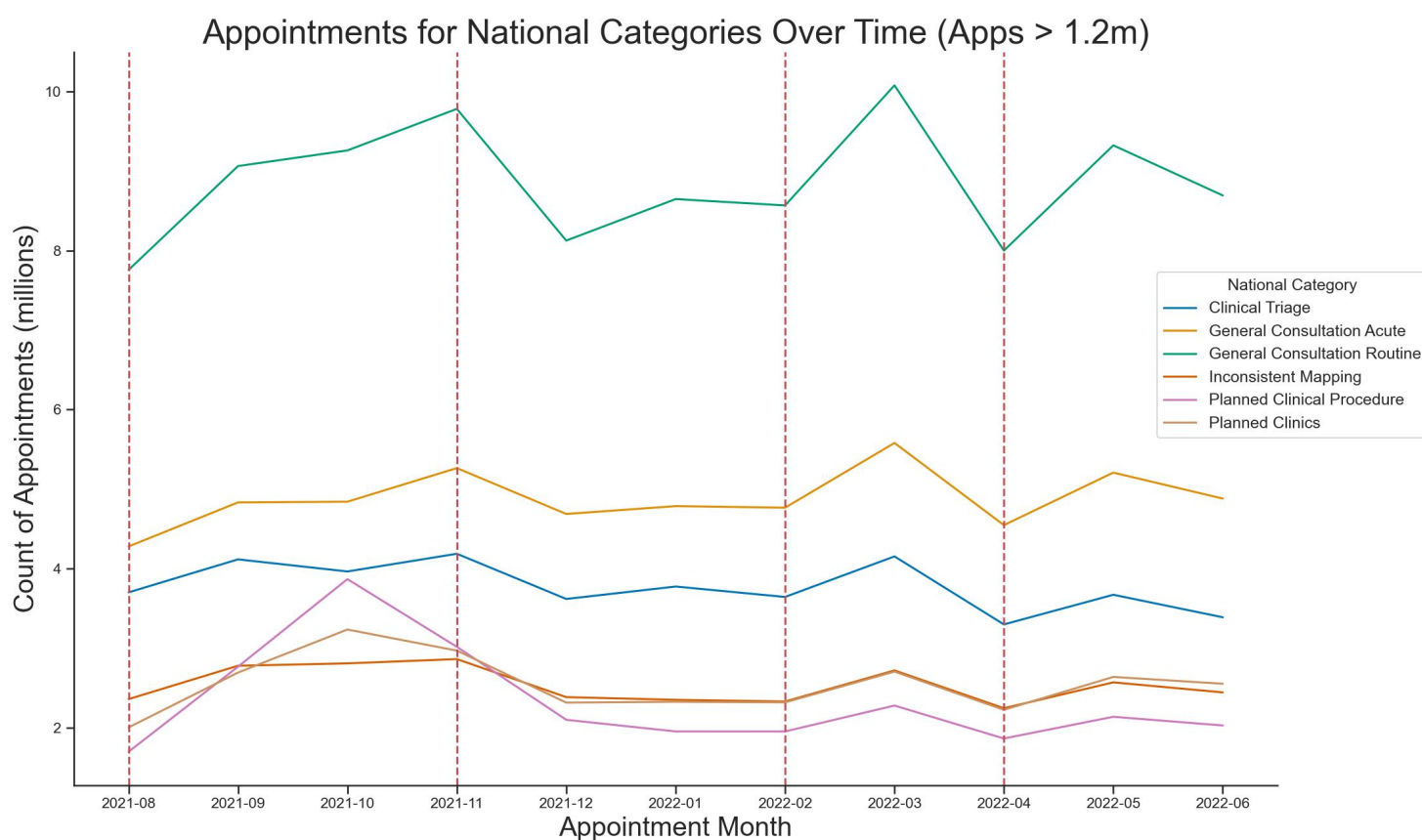
- There is a rise in appointments for most **National Categories**, particularly **General Consultation Routine** and **Planned Clinical Procedure** in the busiest months.
  - "The NHS is always under considerable pressure over the winter period as demand for services tends to increase significantly with the onset of cold weather and flu."
    - NHS Providers
  - Perhaps the decrease in appointment numbers in **December** was due, in part, to the shortages in staff across the NHS over the Christmas 2021 period.

- As the busiest months have been revealed, along with Service Settings, Context Types, and National Categories, I thought it pertinent to investigate Healthcare Provider Types in relation to these trends.

The Change in Healthcare Provider Types Over Time
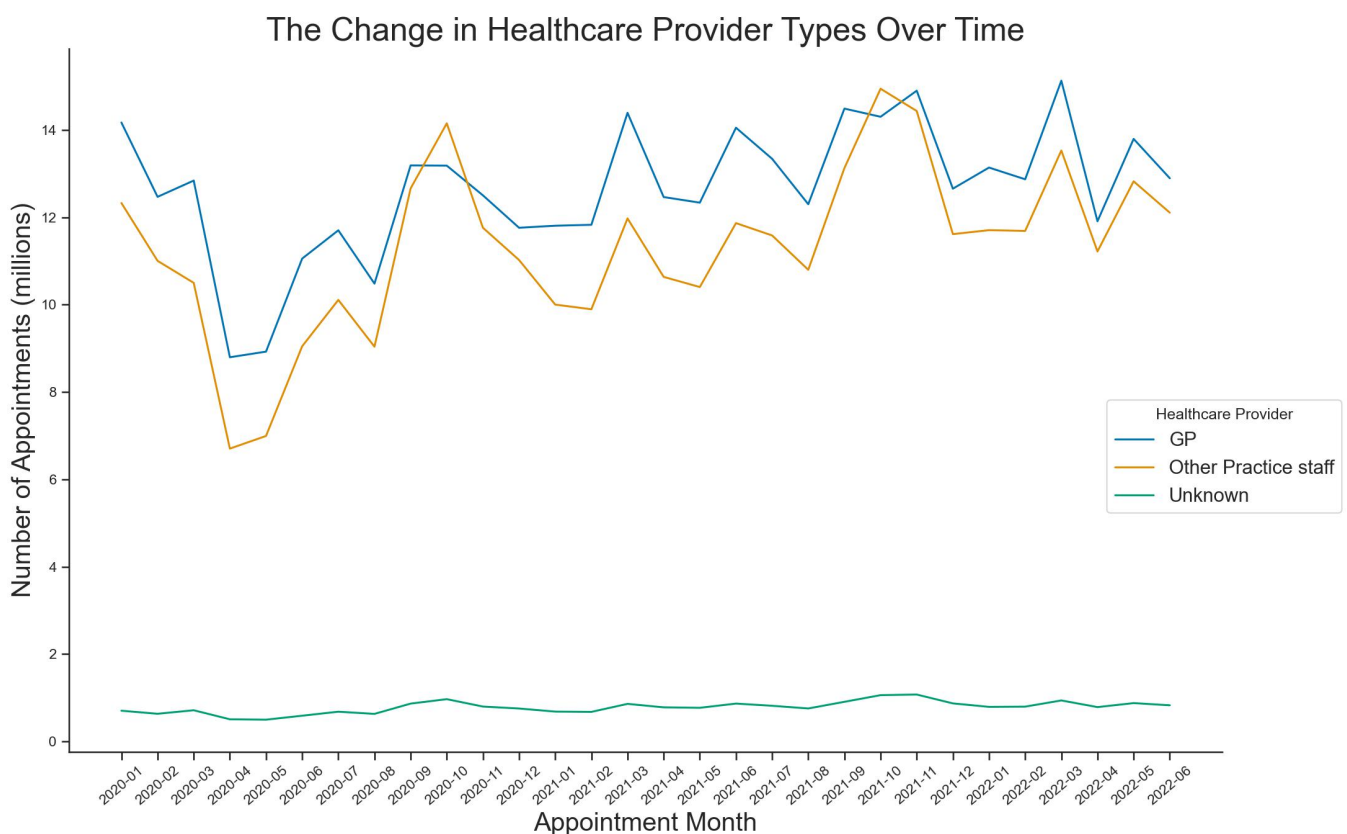
- **GP** and **Other Practice Staff** share a similar trajectory and peak in the busiest periods.
  - It would be useful to access more specific staff data to break the **Other Practice Staff** category into its contained roles and see how in-demand they are during busy periods.

- There was a significant drop in appointments between **January 2020** and **April 2020.**
  - This may be explained by introduction of the first nation-wide lockdown in **March 2020**, which meant that most people couldn't physically attend appointments.
    - This does not necessarily mean that all appointments dropped in number during this time.

The Change in Appointment Modes over Time
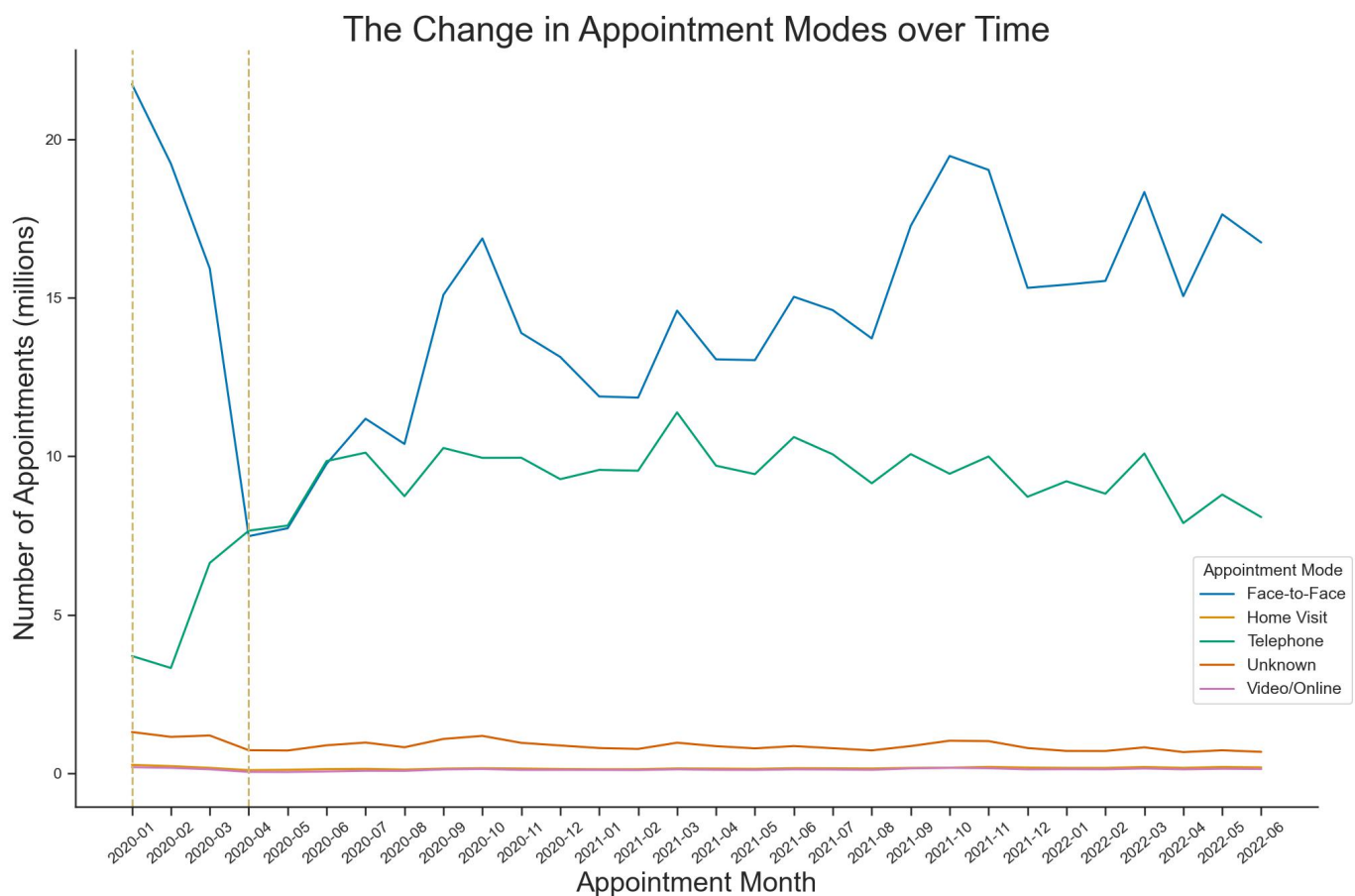
- Patients sought to utilise **Telephone** appointments, as it is the only mode to rise in this period.
  - Was the NHS prepared for this rise in telephone appointments?
    - Perhaps it is worth conducting sentiment analysis regarding patients' satisfaction with the service throughout this period.

- There is a gradual fall in **Telephone** appointments and rise in **Face-to-Face**.
  - This may be a result of the UK vaccine roll-out in late 2020 / early 2021 and the easing of lockdown restrictions.

- During the busiest months (**September - November 2021** and **March 2022**) the number of **Face-to-Face** appointments rose significantly.
  - This is a clear indication of where the NHS' resources are being utilised in busy periods.

- Gaining access to **National Categories** data for **January 2020** onwards would provide beneficial insight into where resources (eg. **Service Setting**) were utilised during this period.
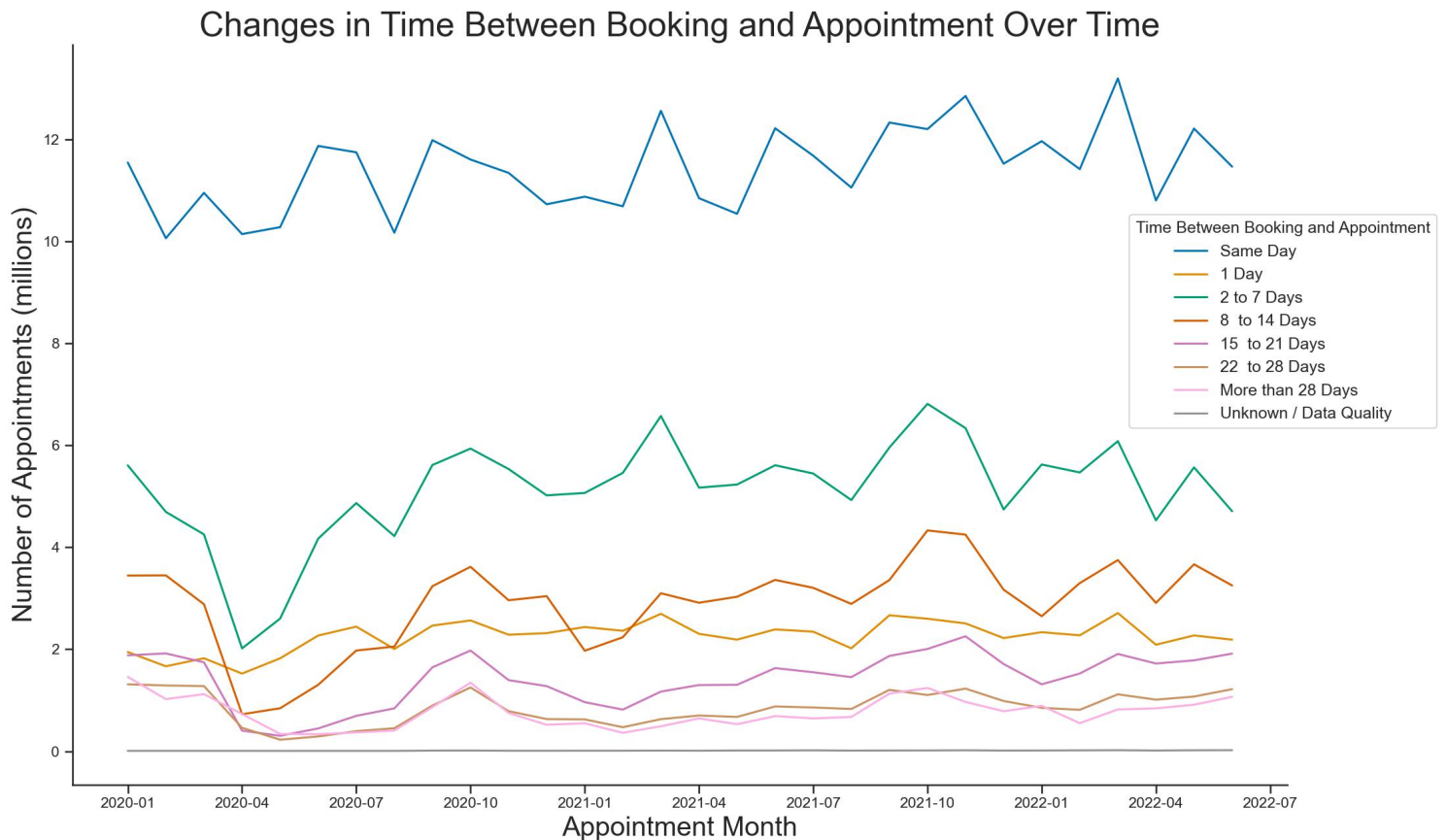
- **March 2020**, **2021**, and **2022** are consistently busy, particularly for **Same Day** appointments.

- Appointments with shorter waiting times rose and fell sharply in line with the busiest periods.
  - It may be useful to compare **appointment status** alongside appointments within the **Same Day** category.

- I did not find sufficient data to accurately determine staff levels in the NHS and whether they were adequate during the period in the dataset.
  - As the project progresses, it would be significantly beneficial to the project goal to access data more directly related to this.

Twitter Data

- Although analysis of the twitter data provided some insight into hashtag trends, because I could not extract time-stamp information from the data set, I could not be sure the insights would be relevant to the project at this time.
  - It would be useful to delve deeper into the twitter data to conduct sentiment analysis, particularly in relation to the NHS app, released in January 2019.
    - It was the most downloaded free app by UK iPhone users in 2021.
    - It encourages users to take up online services, which may help with missing or incorrectly recorded data.