

Background/context of the business scenario

The NHS incurs significant, potentially avoidable, costs when patients miss general practitioner (GP) appointments. Therefore, reducing or eliminating missed appointments would be beneficial financially as well as socially. This analysis will investigate the capacity of the networks and staff involved and the actual utilisation of resources.

Analytical approach

For each of the dataset provided, pandas library was used to parse in the relevant data into dataframes. Sense check was conducted, looking at the shape, null values, first and last few records of the dataframe. After which, frequent saves were conducted on github. For data visualisation, plotly.express was used as the main library given its simplicity in code inputs and clear and interactive plots. Given that the dataset has a time element, line plots were used for most of the analysis to visualise the trends across several time periods. (Refer to the assignments header for more detailed information)

Assignment 3

Questions were asked on the national category data set.

- 1) Conducted sense check of the appointment date format using `.head()`. Changed the date format to `datetime64[ns]` and used `min()`, `max()` to get the earliest and latest date appointments. Used F-string for the output.

For the minimum and maximum dates in the `ad` DataFrame:

```
First date of scheduled appointments: 2021-12-01 00:00:00
Last date of scheduled appointments: 2022-06-30 00:00:00
```

For the minimum and maximum dates in the `nc` DataFrame:

```
First date of scheduled appointments: 2021-08-01 00:00:00
Last date of scheduled appointments: 2022-06-30 00:00:00
```

- 2) The question requires 3 filters: -
 - a. Records from 'NHS North West London ICB - W2U3Z'
 - b. Appointment date before 01 June 2022.
 - c. Appointment date after 01 January 2022.Used pandas `.loc` function and assigned it to a new variable. Used `value_counts()`, `sort_values()` and `.head()` to the new variable to get the most popular service setting. (General Practice with 2104 appointments)

```
General Practice    2104
Name: service_setting, dtype: int64
```

- 3) There were several ways to get the highest number of appointments. One of which was using `groupby` function with the `series.dt.year` and `series.dt.month` and summing up the appointment counts. In addition, used `sort_values`, `reset_index` and `rename` columns for clarity.

	year	month	count_of_appointments
0	2021	11	30405070
1	2021	10	30303834
2	2022	3	29595038
3	2021	9	28522501
4	2022	5	27495508
5	2022	6	25828078
6	2022	1	25635474
7	2022	2	25355260
8	2021	12	25140776
9	2022	4	23913060
10	2021	8	23852171

- 4) Similar to 3), the nc dataframe had an appointment column which can be used for a groupby function to get the sum of count of appointments. `.reset_index()` was used for clarity.

	appointment_month	count_of_appointments
0	2021-08	23852171
1	2021-09	28522501
2	2021-10	30303834
3	2021-11	30405070
4	2021-12	25140776
5	2022-01	25635474
6	2022-02	25355260
7	2022-03	29595038
8	2022-04	23913060
9	2022-05	27495508
10	2022-06	25828078

Assignment 4

The questions require the use of groupby function extensively.

The data type of appointment month was changed to string for easier plotting later.

A separate data set was created for 4.1 and 4.2 each.

4.1

Using the national category dataframe, the data was further grouped into service settings, context type, national category, months and counting the number of appointments. Column index was reset. Using plotly, line graphs were plotted to find trends.

Noted that majority of the appointments stems from general practice, care related encounter and general consultation.

Out[58]:

	appointment_month	service_setting	count_of_appointments
0	2021-08	Extended Access Provision	160927
1	2021-08	General Practice	21575852
2	2021-08	Other	449101
3	2021-08	Primary Care Network	432448
4	2021-08	Unmapped	1233843

Out[58]:

	appointment_month	context_type	count_of_appointments
0	2021-08	Care Related Encounter	20255235
1	2021-08	Inconsistent Mapping	2363093
2	2021-08	Unmapped	1233843
3	2021-09	Care Related Encounter	24404251
4	2021-09	Inconsistent Mapping	2782135

Out[58]:

	appointment_month	national_category	count_of_appointments
0	2021-08	Care Home Needs Assessment & Personalised Care...	29676
1	2021-08	Care Home Visit	47583
2	2021-08	Clinical Triage	3704207
3	2021-08	General Consultation Acute	4280920
4	2021-08	General Consultation Routine	7756045

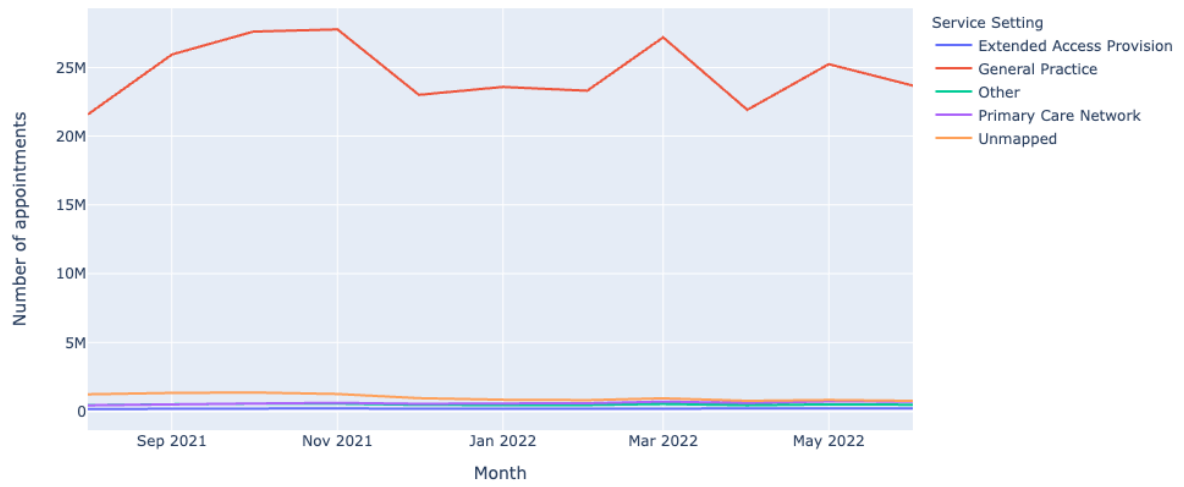
For the month of August 2021,

About 50% of national category appointments related to General consultation routine.

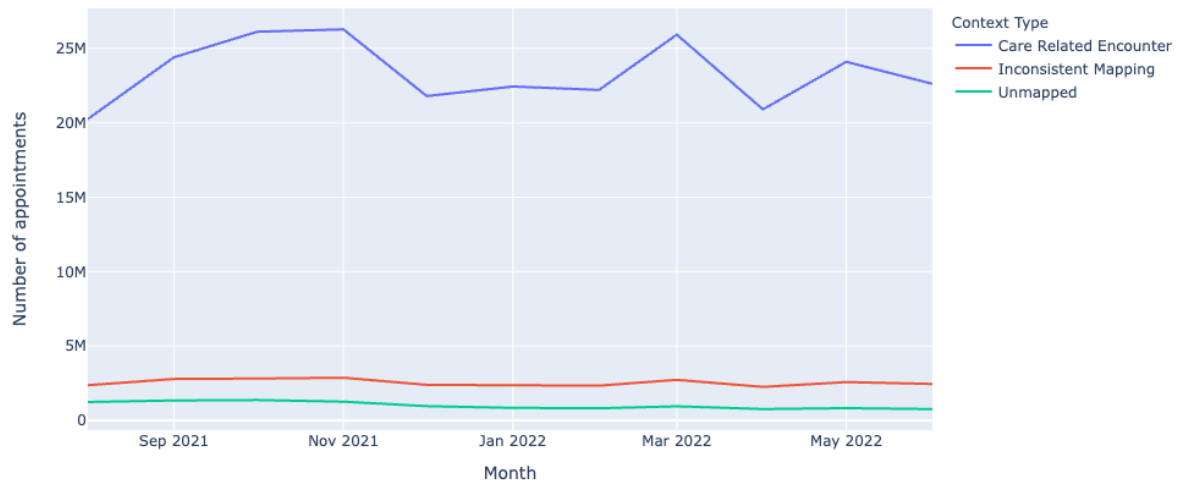
Over 90% of context type were related to care related encounters.

Over 90% of service settings were related to general practice.

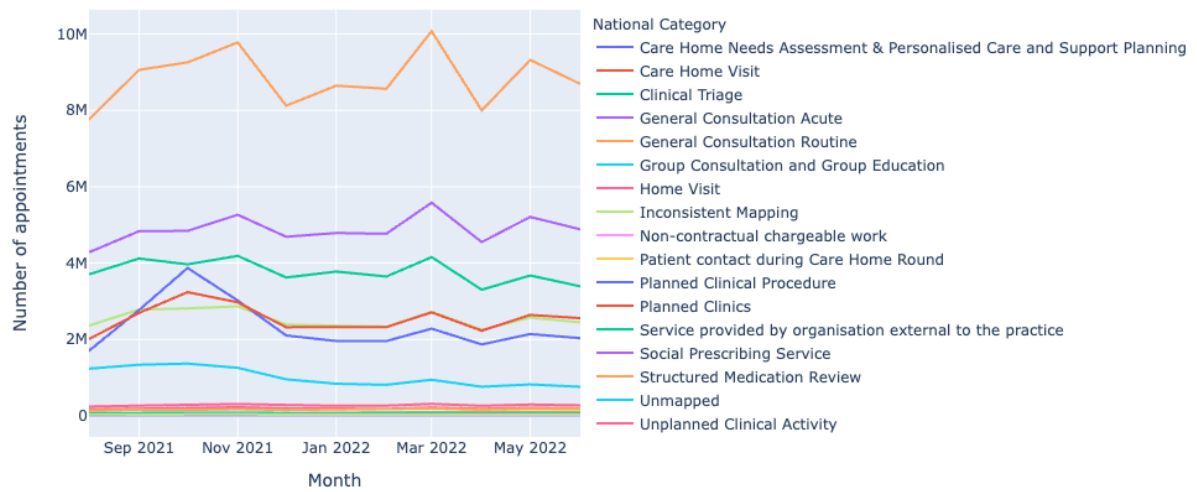
Total appointment across months and service settings



Total appointment across months and context types



Total appointment across months and national categories



4.2

Similar to 4.1, the dataframe was grouped by appointment date, month, service setting and counting the number of appointments. Plotly was used to visualize the findings.

	appointment_date	appointment_month	service_setting	count_of_appointments
0	2021-08-01	2021-08	Extended Access Provision	438
1	2021-08-01	2021-08	General Practice	3411
2	2021-08-01	2021-08	Other	401
3	2021-08-01	2021-08	Primary Care Network	323
4	2021-08-01	2021-08	Unmapped	1054
...
150	2021-08-31	2021-08	Extended Access Provision	8281
151	2021-08-31	2021-08	General Practice	1125584
152	2021-08-31	2021-08	Other	23181
153	2021-08-31	2021-08	Primary Care Network	21789
154	2021-08-31	2021-08	Unmapped	62380

155 rows × 4 columns

Out[109]:

	appointment_date	appointment_month	service_setting	count_of_appointments
305	2021-10-01	2021-10	Extended Access Provision	7107
306	2021-10-01	2021-10	General Practice	1130646
307	2021-10-01	2021-10	Other	22639
308	2021-10-01	2021-10	Primary Care Network	22974
309	2021-10-01	2021-10	Unmapped	56201
...
455	2021-10-31	2021-10	Extended Access Provision	1113
456	2021-10-31	2021-10	General Practice	17449
457	2021-10-31	2021-10	Other	1147
458	2021-10-31	2021-10	Primary Care Network	1200
459	2021-10-31	2021-10	Unmapped	2863

155 rows × 4 columns

Out[109]:

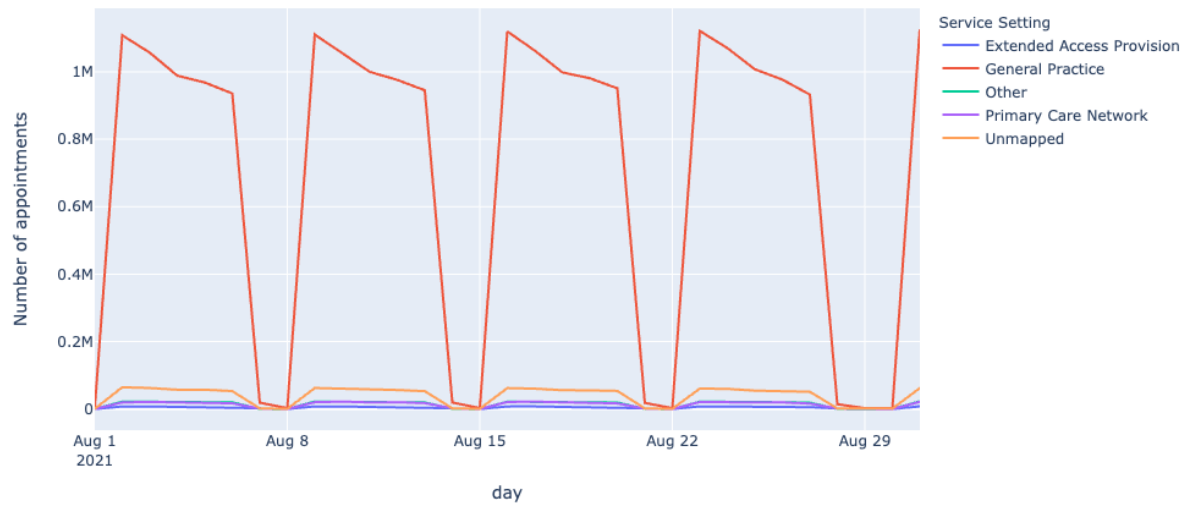
	appointment_date	appointment_month	service_setting	count_of_appointments
765	2022-01-01	2022-01	Extended Access Provision	246
766	2022-01-01	2022-01	General Practice	3578
767	2022-01-01	2022-01	Other	280
768	2022-01-01	2022-01	Primary Care Network	215
769	2022-01-01	2022-01	Unmapped	572
...
915	2022-01-31	2022-01	Extended Access Provision	9962
916	2022-01-31	2022-01	General Practice	1305082
917	2022-01-31	2022-01	Other	25034
918	2022-01-31	2022-01	Primary Care Network	30509
919	2022-01-31	2022-01	Unmapped	43240

155 rows × 4 columns

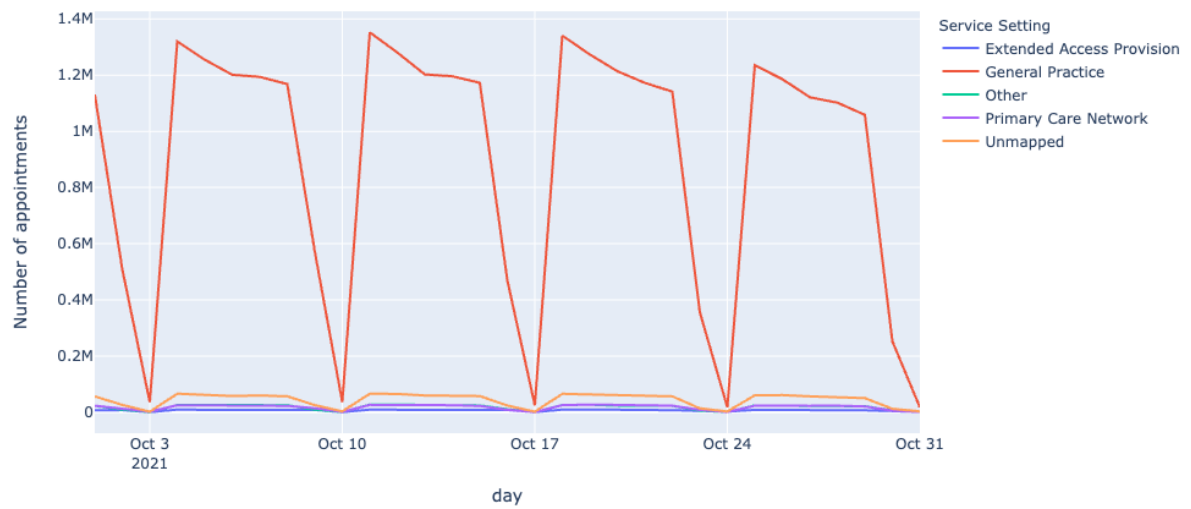
Out[109]:

	appointment_date	appointment_month	service_setting	count_of_appointments
1215	2022-04-01	2022-04	Extended Access Provision	8011
1216	2022-04-01	2022-04	General Practice	1085037
1217	2022-04-01	2022-04	Other	21077
1218	2022-04-01	2022-04	Primary Care Network	26493
1219	2022-04-01	2022-04	Unmapped	36836
...
1360	2022-04-30	2022-04	Extended Access Provision	3416
1361	2022-04-30	2022-04	General Practice	23675
1362	2022-04-30	2022-04	Other	1148
1363	2022-04-30	2022-04	Primary Care Network	3150

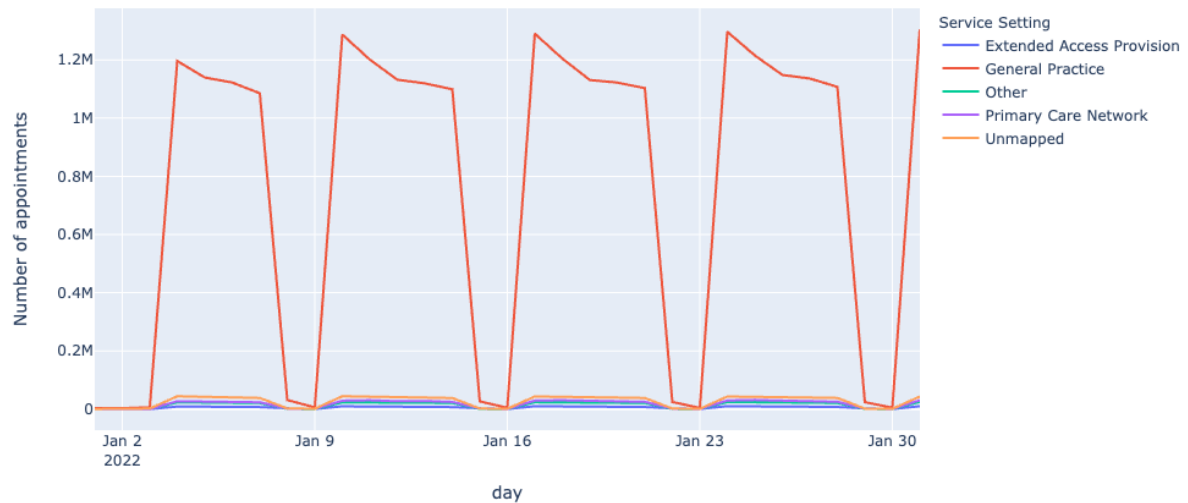
Total appointment across summer and service settings



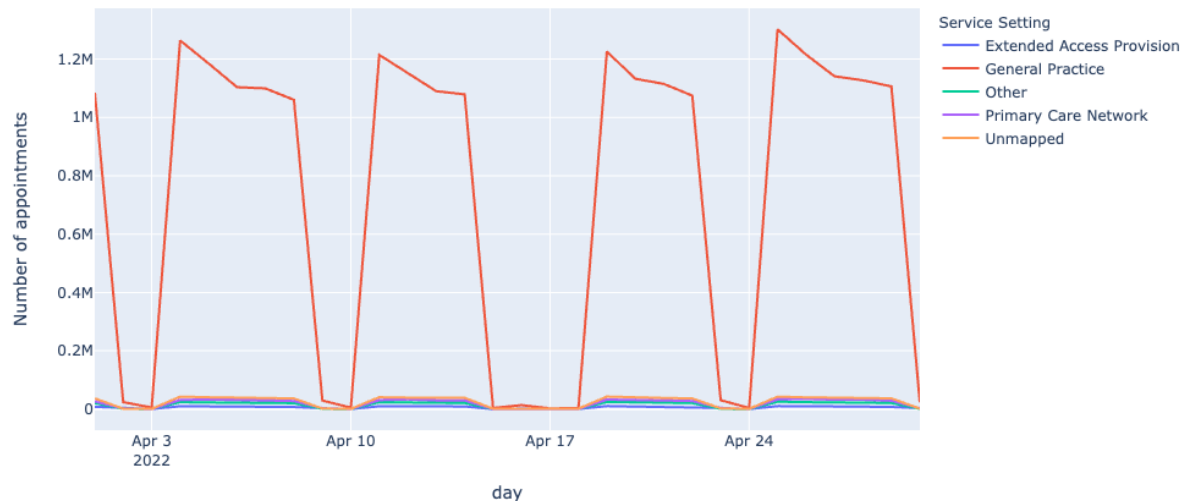
Total appointment across autumn and service settings



Total appointment across winter and service settings



Total appointment across spring and service settings



Findings

Seasonal visualisations show an interesting trend where there is a low number of appointments every Sunday and a high number of appointments every Monday for general practice. In addition, the number of appointments generally decreases from Monday to Sunday. This trend for general practice is consistent across all seasons.

This can plausibly be because most patients' well-being was affected during working (Monday to Friday) and non-working days. (Saturday – Sunday)

NHS may want to look into the relationship between working days and medical appointments.

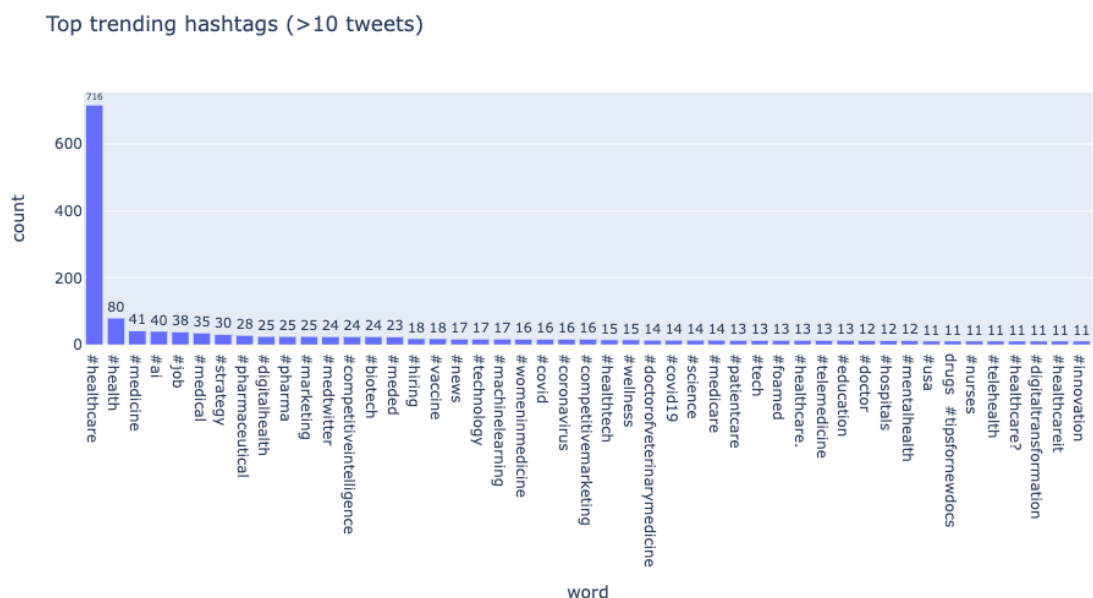
For the remaining service settings, the low number of appointments is consistent across all seasons and months.

Assignment 5

From the tweets dataset, the data contains retweets and favourite tweets; essentially a repetition of the original tweet. Hence, it is not useful to only look at both columns when we are trying to draw insights. Based on the value counts of both columns, more than half of those counts were empty. It will be useful to focus on the original tweet instead.

To extract the hashtag from the `tweet_full_text` column, it is best to iterate through the column to extract out substrings containing hashtag and its count and subsequently append it to an empty list.

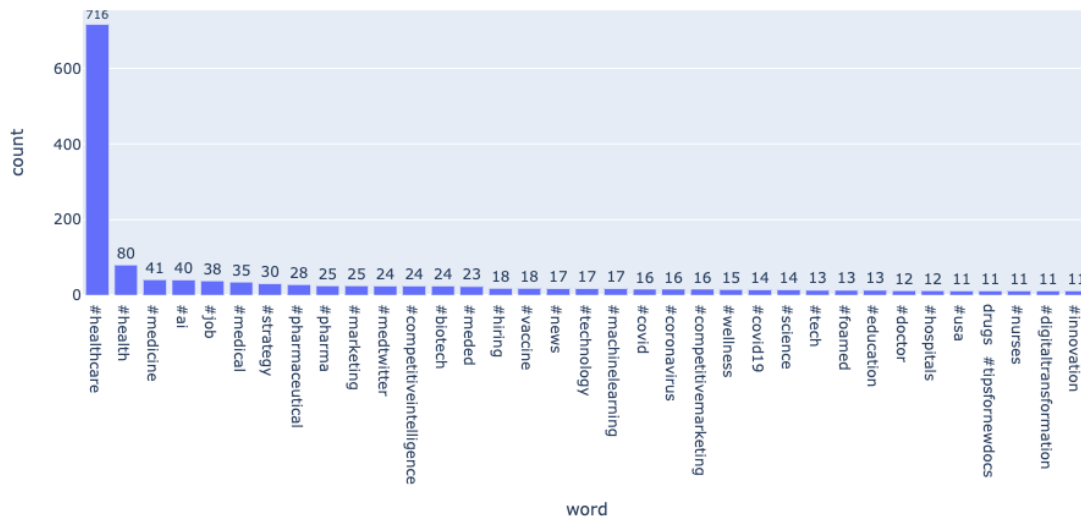
To visualize the data, the hashtag list will be parsed into a dataframe, resetting the index and renaming the columns for consistency and clarity. The count column will be adjusted to an integer datatype. Lastly, only hashtags with more than 10 tweets will be visualized using plotly library.



Based on the barplot above, noted that there were a few overrepresented hashtags. Those hashtags included words like medicine, care and health. To remove those hashtags while leaving a main hashtag. (healthcare, health, medicine) .Drop() method was used after filtering the dataset with .str.contains().

The final cleaned dataset was visualized as follows.

Top trending hashtags (>10 tweets)



Findings

Based on current trending hashtags on twitter, the top 3 hashtags were all related to Health. (health, healthcare and medicine)

Tweeting can be a powerful tool to raise awareness about clinical appointments, engage with patients and develop leads. It also increases exposure towards the medical industry. Missing medical appointments delays treatment for other patients in need, creates stress on the medical staff and capacity and ultimately increases mortality rate. The NHS can plausibly set up a verified business twitter account to raise awareness of the cost of missing medical appointments. In addition, should there be any vital information to alert the people, twitter represents one of the fastest social media platforms to reach the masses.

Assignment 6

1. Should the NHS start looking at increasing staff levels?

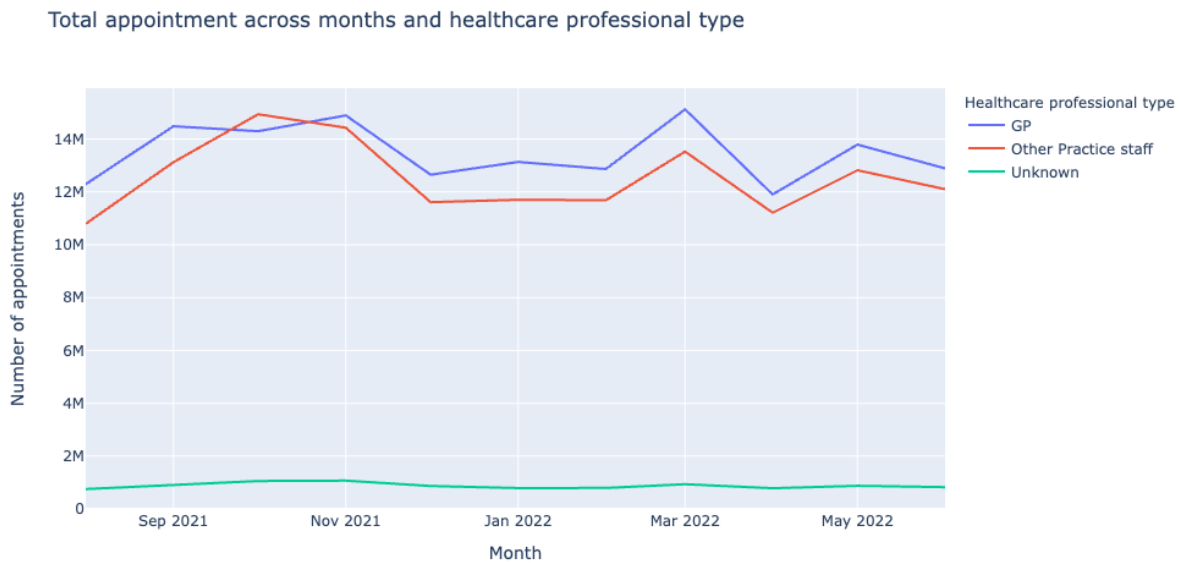
	appointment_month	count_of_appointments	daily_utilisation
0	2021-08	23852171	66.3
1	2021-09	28522501	79.2
2	2021-10	30303834	84.2
3	2021-11	30405070	84.5
4	2021-12	25140776	69.8
5	2022-01	25635474	71.2
6	2022-02	25355260	70.4
7	2022-03	29595038	82.2
8	2022-04	23913060	66.4
9	2022-05	27495508	76.4
10	2022-06	25828078	71.7

Based on the data above, the highest average daily utilization capacity rate reached 84.5% during October 2021(30405070 appointments)

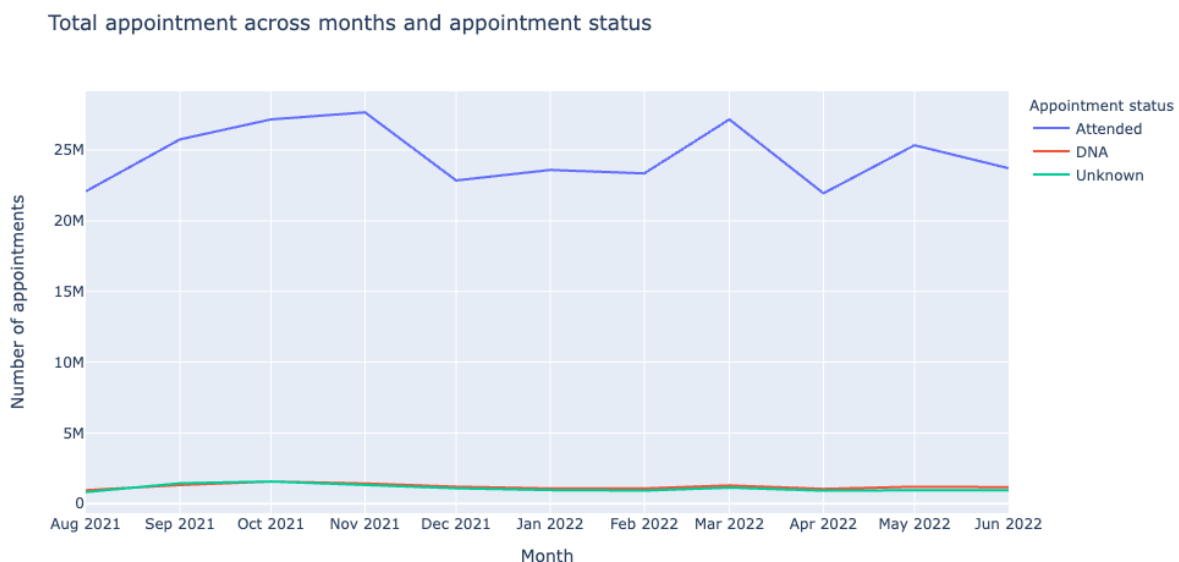
while the average monthly capacity utilization rate tends to hover between 65% to 85%. The NHS still have a buffer of about 15% capacity utilisation and should not investigate increasing staff levels.

2. How do the healthcare professional types differ over time?

There is not much difference between healthcare professional types. Except for October 2021, most known professions apart follows the same trend throughout the same time period. Most of the patients made appointment for general consultations. (GP)



3. Are there significant changes in whether or not visits are attended?

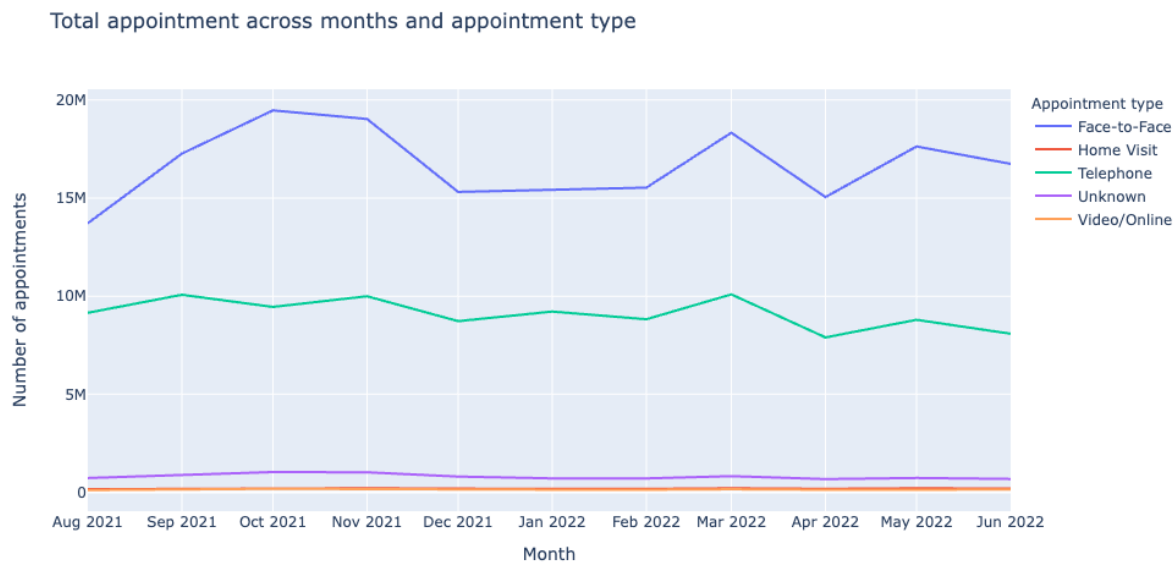


No significant changes were found.

Average appointment attended ranged between 20 – 30 million.

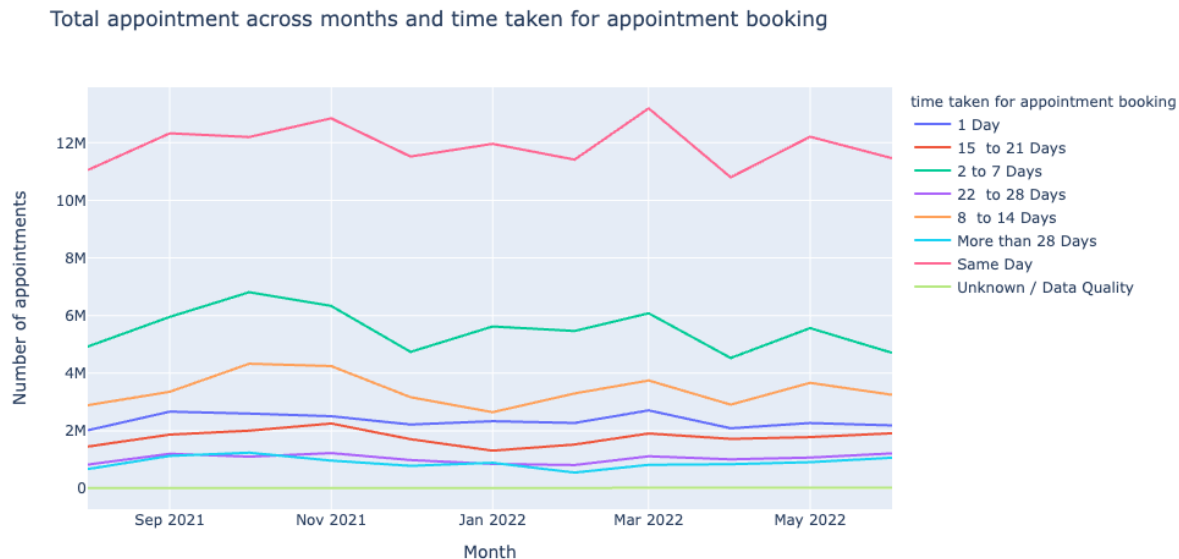
There were very low number of patients not attending appointments. (< 1.57 million per month)

4. Are there changes in terms of appointment type and the busiest months?



No significant changes were found. Majority of patients preferred Face-to-Face and Telephone consultations. NHS can look into reallocating GP appointments to teleconsultations.

5. Are there any trends in time between booking and appointment?



There were a significant number of patients making appointments on the same day, followed by appointments made within a week, 2 weeks and 1 day in advance.

6. How do the various service settings compare?

Boxplots were used for the visualization. However, noted that it is not easy to visualize given the range of appointment values at which the different service settings are operating at, particularly for General Practice. Hence, isolating

general practice and looking at the remaining service settings would provide a slightly clearer visualization.

Without general practice appointments, we can see that most of the other service settings generally follows the same trend across the months. However, there were a number of unmapped service setting appointments. The NHS can look into reducing these numbers by investigating further into the reasons why these appointments are unmapped. In doing so, will allow NHS to be more efficient in allocating resources.

