JULY 30, 2023

LSE Data Analytics Online Career Accelerator

Assignment: Diagnostic Analysis using Python

ERDI OZYURT

1. Introduction

1.1. Background

As part of the course, analyse a data set to help the publicly funded healthcare system in England, NHS to understand the reasons for missed appointments to utilize Python to explore the available data, create visualizations to identify trends, and extract meaningful insights to inform decision-making.

According to the NHS, each appointment costs an average of £30, putting the total cost to the NHS at more than £ 216 million pounds on top of the disruption for staff and fellow patients. More than 15 million general practice appointments are being wasted each year because patients do not turn up and fail to warn surgeries that they will not be attending. Of these, around 7.2 million are with busy family doctors, which adds up to more than 1.2 million GP hours wasted each year – the equivalent of over 600 GPs working full time for a year.

To complete this analysis, we will be analysing four datafiles,

- actual_duration.csv,
- appointments_regional.csv,
- national_categories.xlsx,
- tweets.csv,

for data exploration and analysis, we will be using Python and Jupiter Notebooks.

1.2. Objectives

At this project stage, the two main questions posed by the NHS are:

- Has there been adequate staff and capacity in the networks?
- What was the actual utilization of resources?

2. Analytical approach

2.1. Data Exploration

All the data exploration was completed using Pandas, Seaborn, Matplotlib, and NumPy libraries.

For ease of manipulation, I imported the four datasets provided into a Jupyter notebook and converted them into data frames. I utilized the pandas python library to aid me in importing and exploring the datasets, and seaborn and matplotlib libraries for effective visualizations.

In order to make the datasets more manageable when exploring the data, I employed the group by and sort_values functionality. This provided direction for future analysis by highlighting the most relevant criteria to the business problem. Appointments_regional was the only dataset to have missed appointments data, so was considered the best source to be investigated, despite the date being by month.

In order to familiarise ourselves with the data we will determine the following:

1. How many locations are there in the data set?

Analysing the national categories data set we saw that the data refers to 106 different locations.

2. What are the five locations with the highest number of records?

Below we can see the top five locations with the most records. It can be seen that NHS

North West London ICB has the highest number of records.

```
In [38]:  1 # Determine the top five locations based on record count.
          2 nc_altern = nc['sub_icb_location_name'].value_counts()
          3 nc_altern.head(5)
```

```
Out[38]: NHS North West London ICB - W2U3Z              13007
         NHS Kent and Medway ICB - 91Q                 12637
         NHS Devon ICB - 15N                           12526
         NHS Hampshire and Isle Of Wight ICB - D9Y0V   12171
         NHS North East London ICB - A3A8R             11837
         Name: sub_icb_location_name, dtype: int64
```

3. How many service settings, context types, national categories, and appointment statuses are there?

Category	Count of types
Service Settings	5
Context Types	3
National Categories	18
Appointment Statuses	3

2.2. Data Analysis

2.2.1. Introduction

In this chapter, we will perform data wrangling and manipulation to find out for answers to general questions that will help us better aspects to understand raw the data and create visualizations to identify possible trends.

2.2.2. Initial Data Analysis

In this chapter, we will search for answers to general questions that will help you better understand the data.

- National Categories dataset (as nc) date range: 2021-08-01 2022-06-30.
- Actual Duration dataset(as ad) 2021-12-01 2022-06-30 .

Overall data General Practice was the most popular service setting for the dataset.

And as it has been asked for specific location, GP was the most popular service setting for NHS North West London (the location with the highest number of records) from 1 January to 1 June 2022.

In [51]:
```
#group by function.
df4 = df3.groupby(['service_setting'])['count_of_appointments'].sum()
NHSNWL = df4.sort_values(ascending=False)
NHSNWL
```

Out[51]:
```
service_setting
General Practice             4804239
Unmapped                      391106
Other                         152897
Primary Care Network          109840
Extended Access Provision      98159
Name: count_of_appointments, dtype: int64
```

The months of October and November 2021 had the highest number of appointments during the period.

In [53]:
```
# Total number of records per month.
Total_pm = nc.groupby(['appointment_month'])['count_of_appointments'].sum()
Total_pm.sort_values(ascending=False)
```

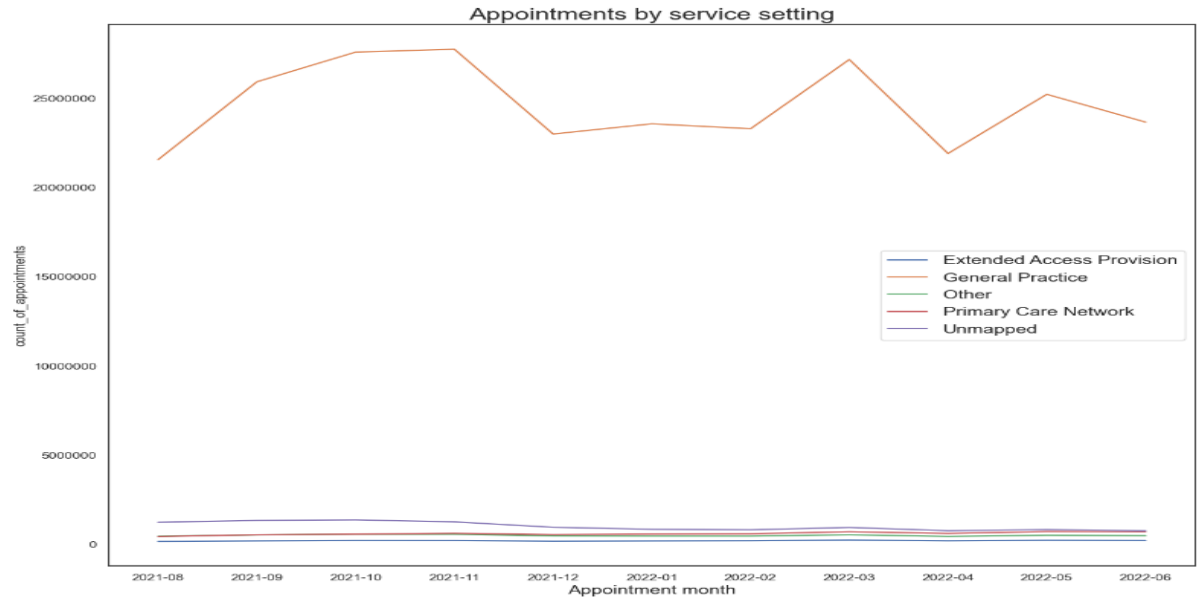
Out[53]:
```
appointment_month
2021-11    30405070
2021-10    30303834
2022-03    29595038
2021-09    28522501
2022-05    27495508
2022-06    25828078
2022-01    25635474
2022-02    25355260
2021-12    25140776
2022-04    23913060
2021-08    23852171
Name: count_of_appointments, dtype: int64
```

2.2.3. Monthly and Seasonal Trends

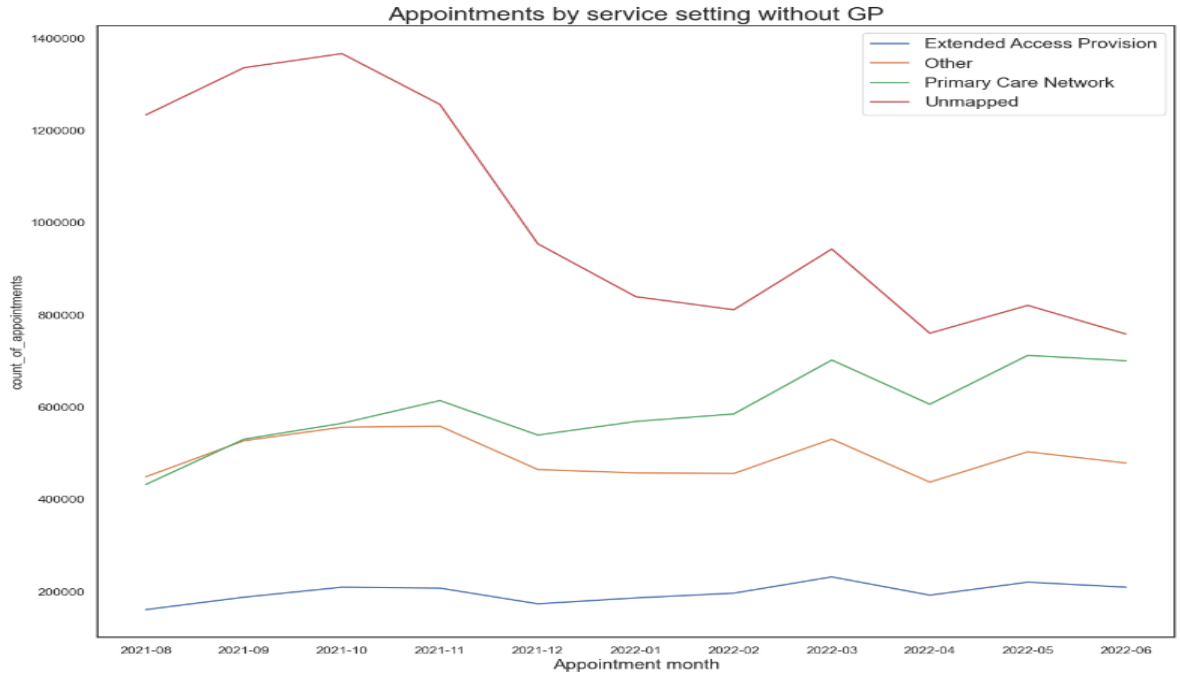
In this chapter, we will find out what monthly and seasonal trends are evident, based on the number of appointments for service settings, context types, and national categories.

Number of appointments per month:

- *Service Settings*

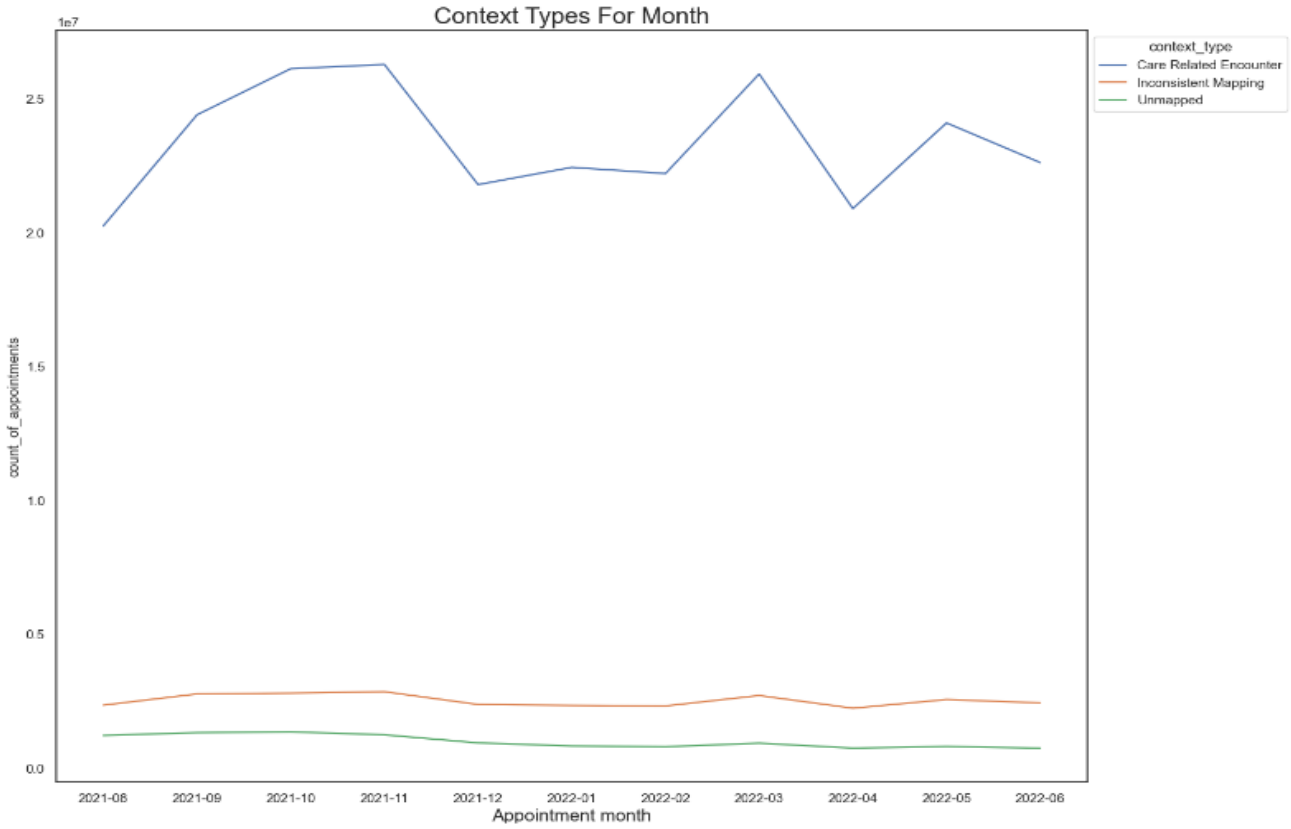


As we all could agree, GP was dominating the all chart and view and we have created another chart excluding GP to see other service setting more visible.



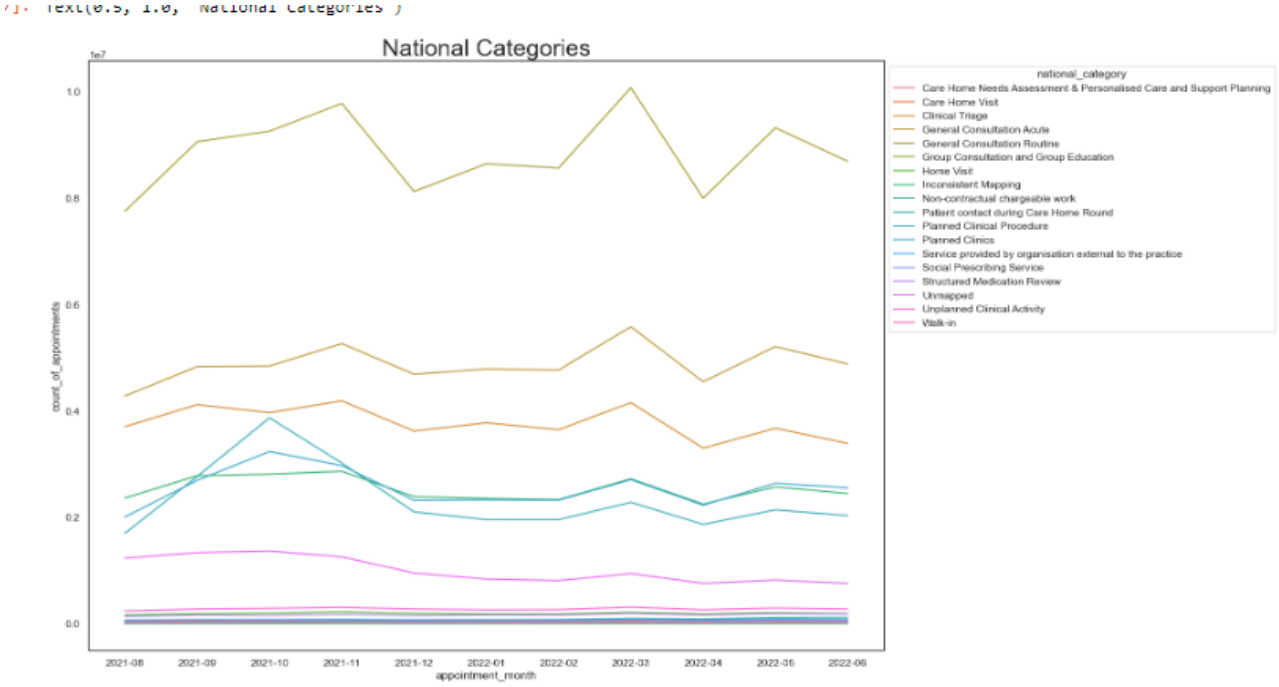
- *Context Types*

The below fig shows that the Care Related Encounter is by far the context type with the highest total number of appointments. This also has peaks in October 2021 and March 2022.



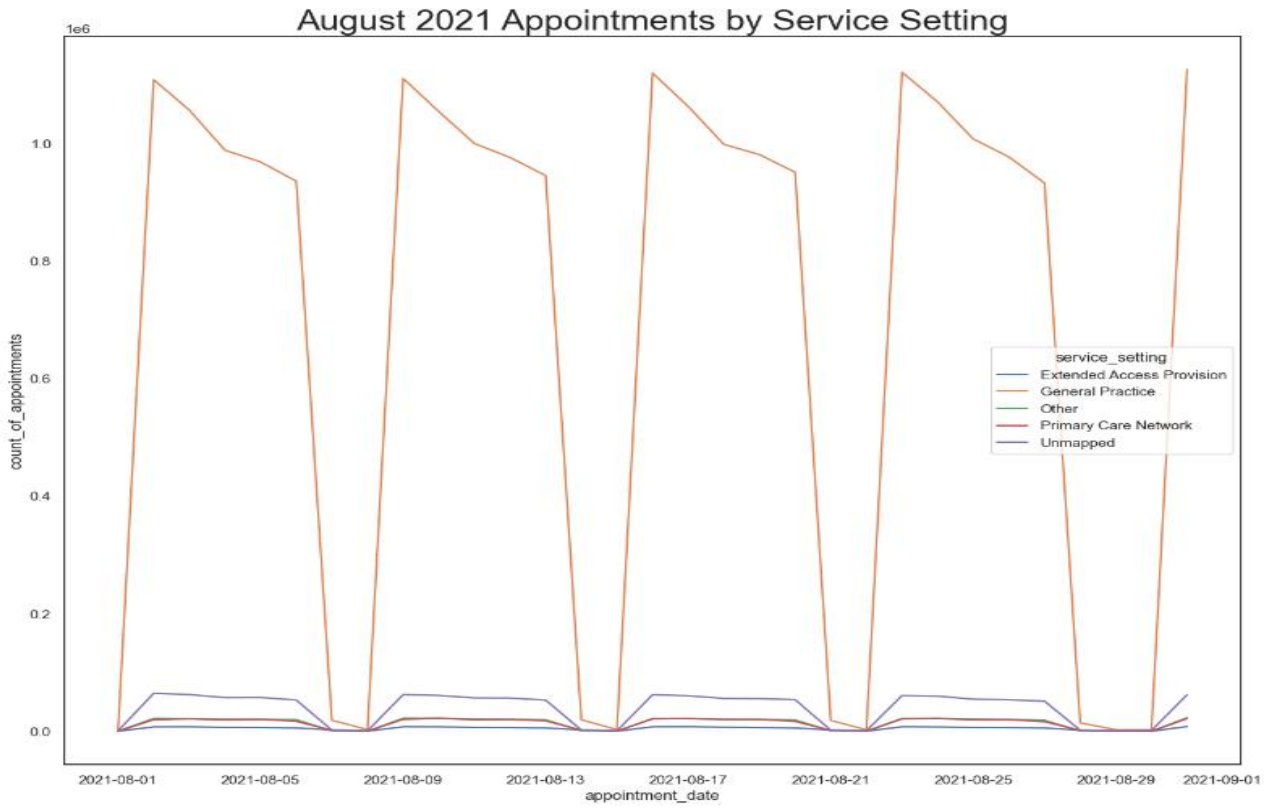
- *National Categories*

The below fig shows that the General Consultation Routine is by far the national category with the highest total number of appointments. This also has peaks in October 2021 and March 2022.

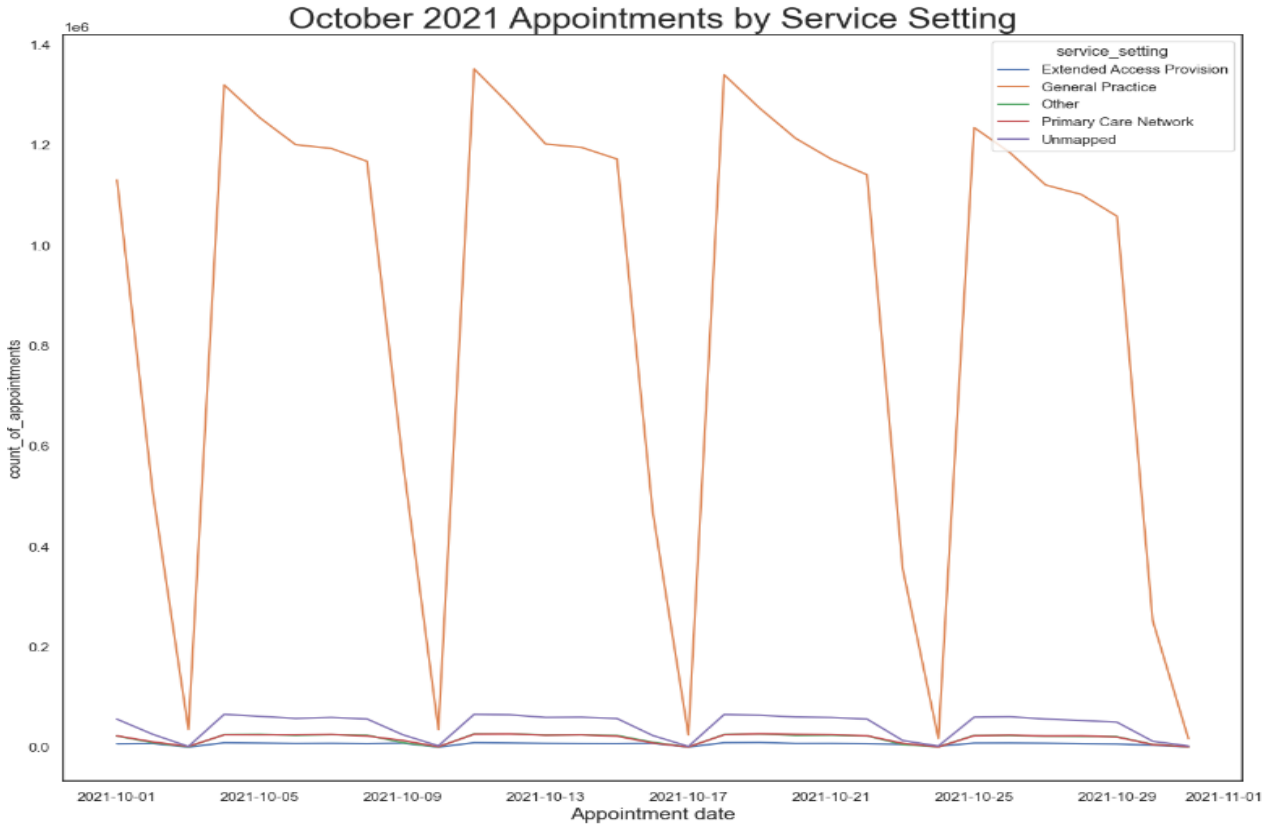


Number of appointments for service setting per season:

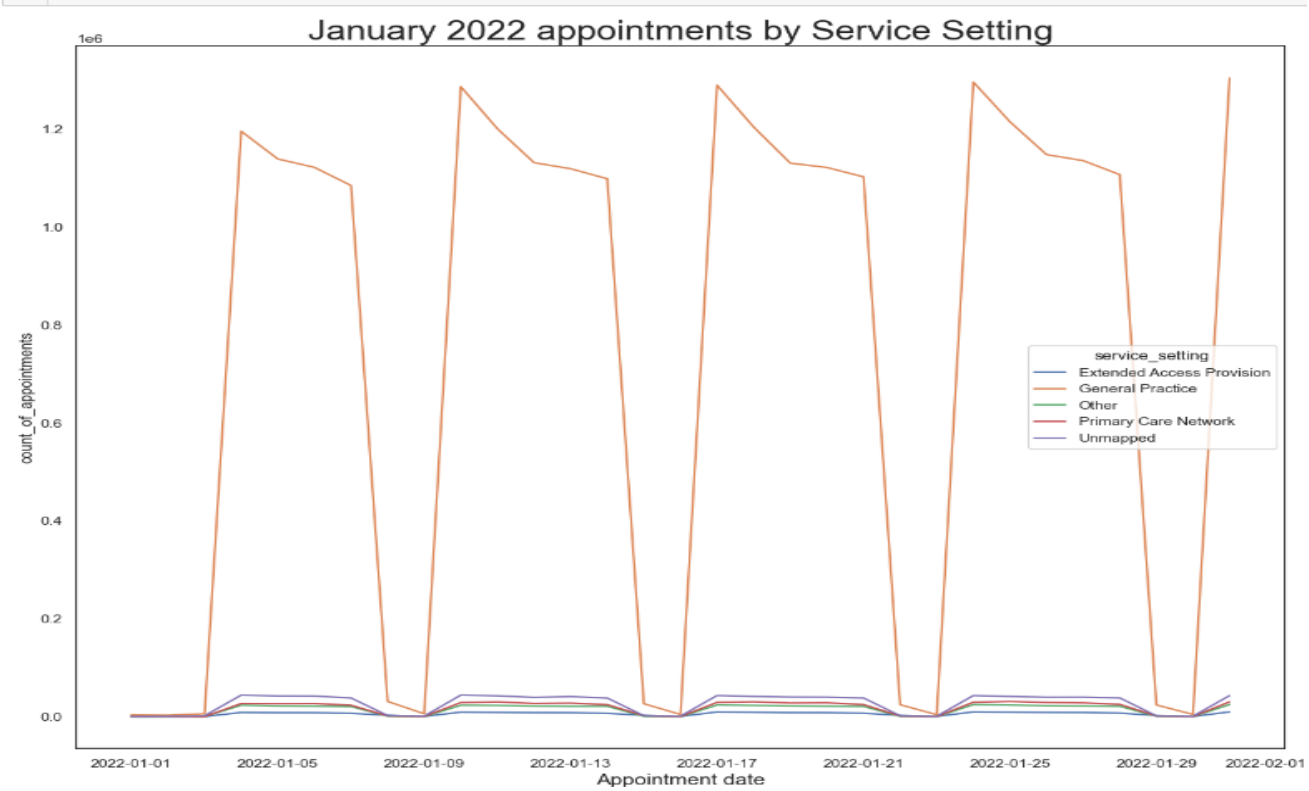
- *Summer (August 2021)*



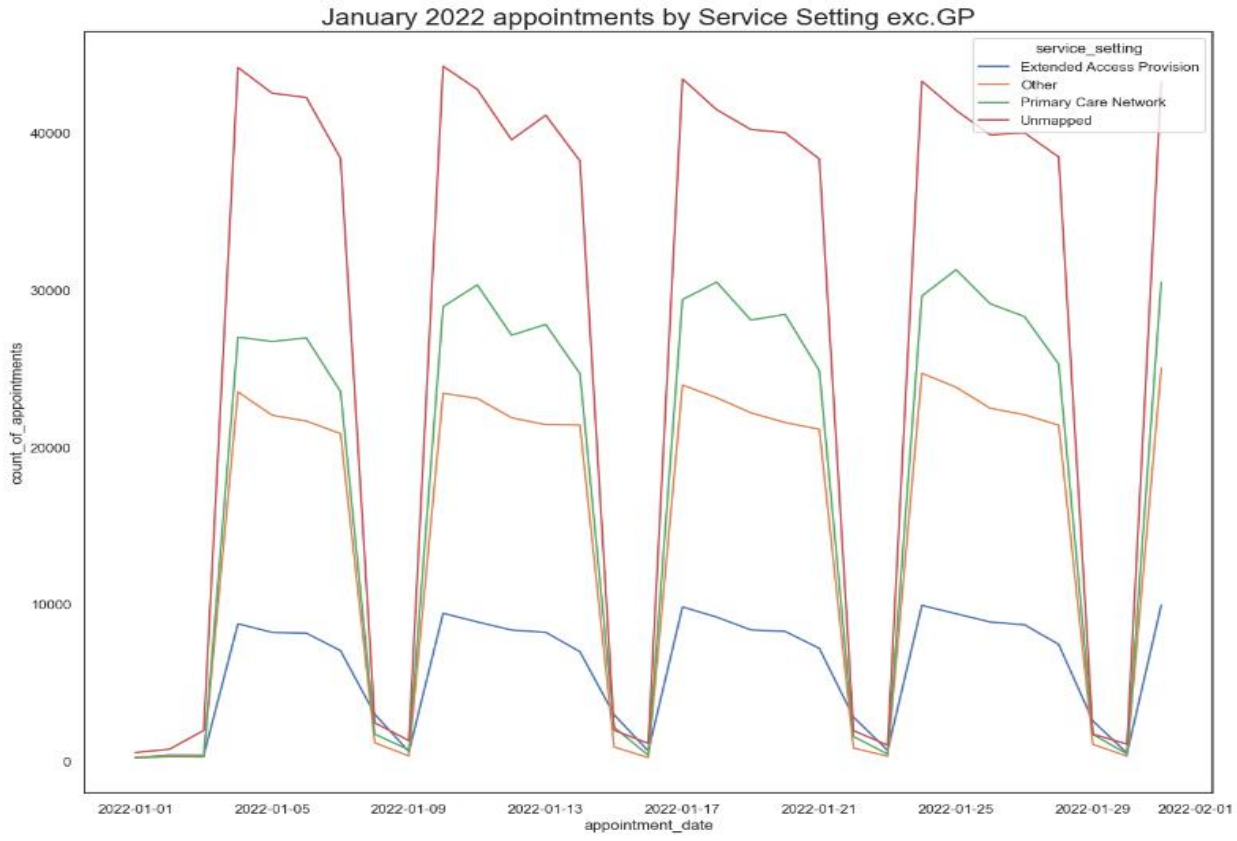
- *Autumn (October 2021)*



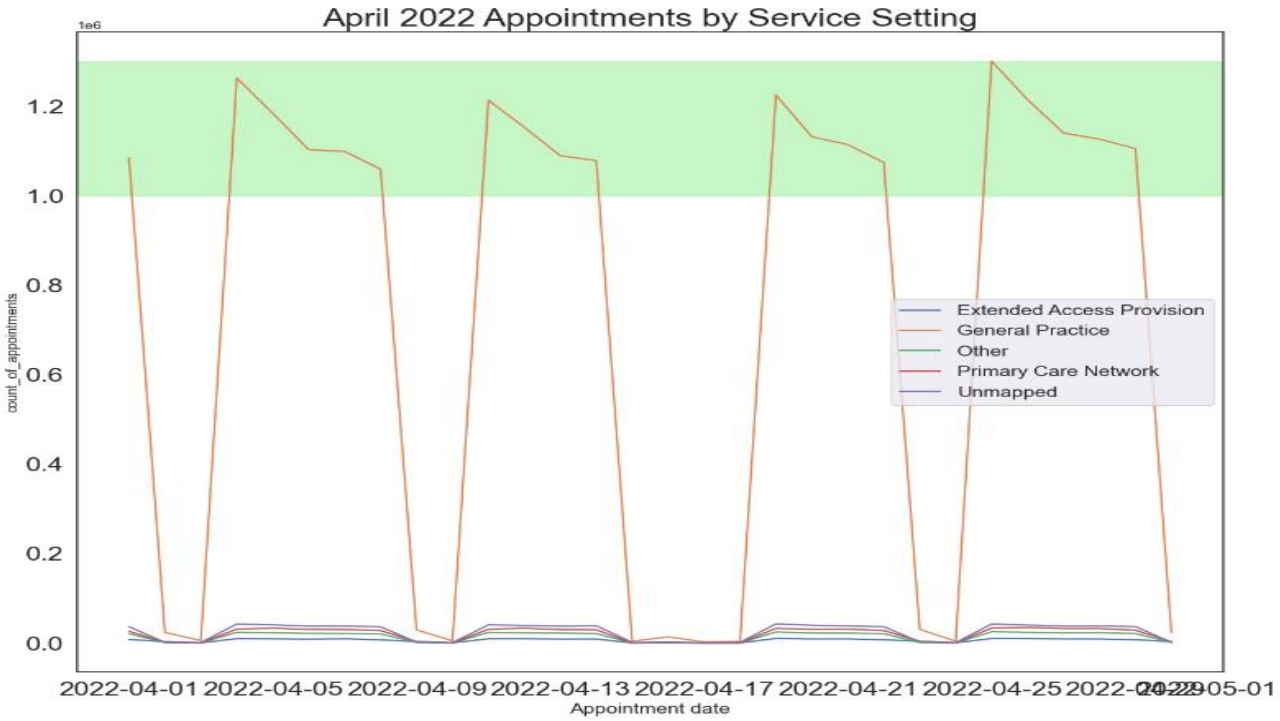
• *Winter (January 2022)*



As we can see all charts have almost the same pattern in terms of numbers and service settings. January the popular season of healthcare, I would like to execute the GP from the chart to examine other settings clearly.



Spring (April 2022)



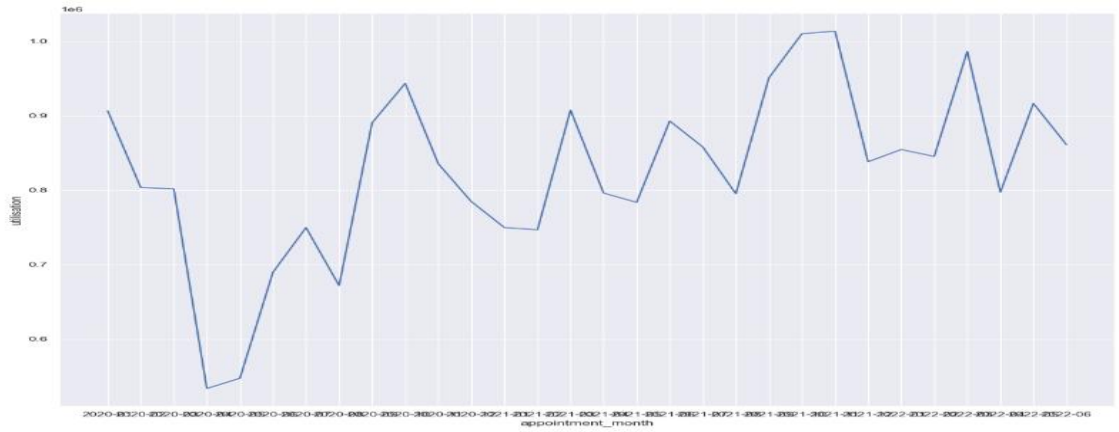
In general Autumn is the season with the highest number of appointments and Summer the lowest. It looks like there is a trend for the appointments to have a peak at the beginning of each week and then start decreasing towards the weekend, it can be seen as well that the appointments go close to zero during the weekends. Knowing that appointment numbers spike on Mondays and drop at weekends suggests that more weekend resource could be required.

3. Findings

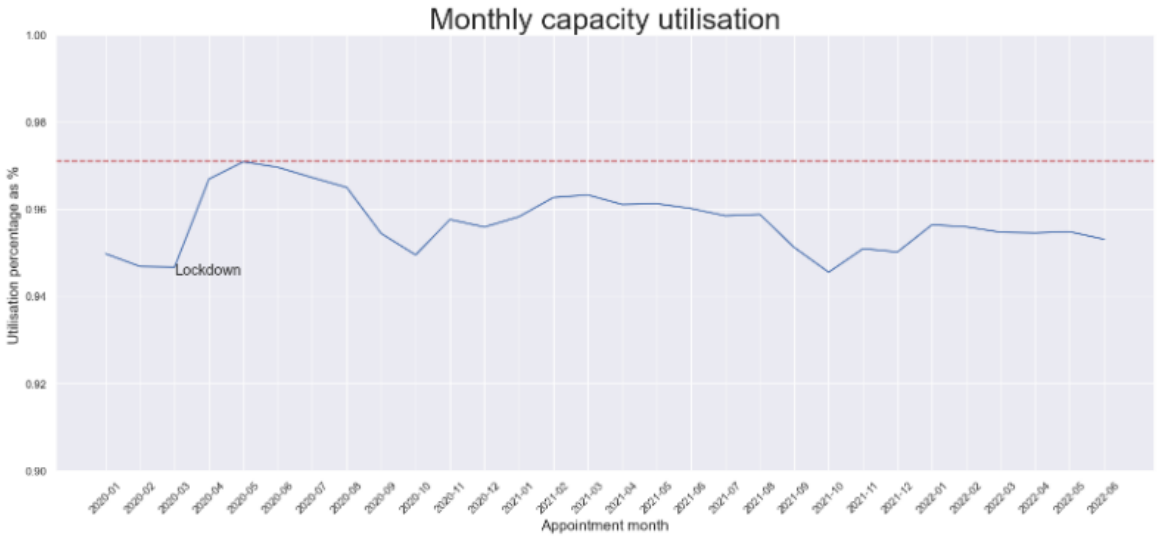
3.1. Patterns, Trends, Insights

- Should the NHS start looking at increasing staff levels?

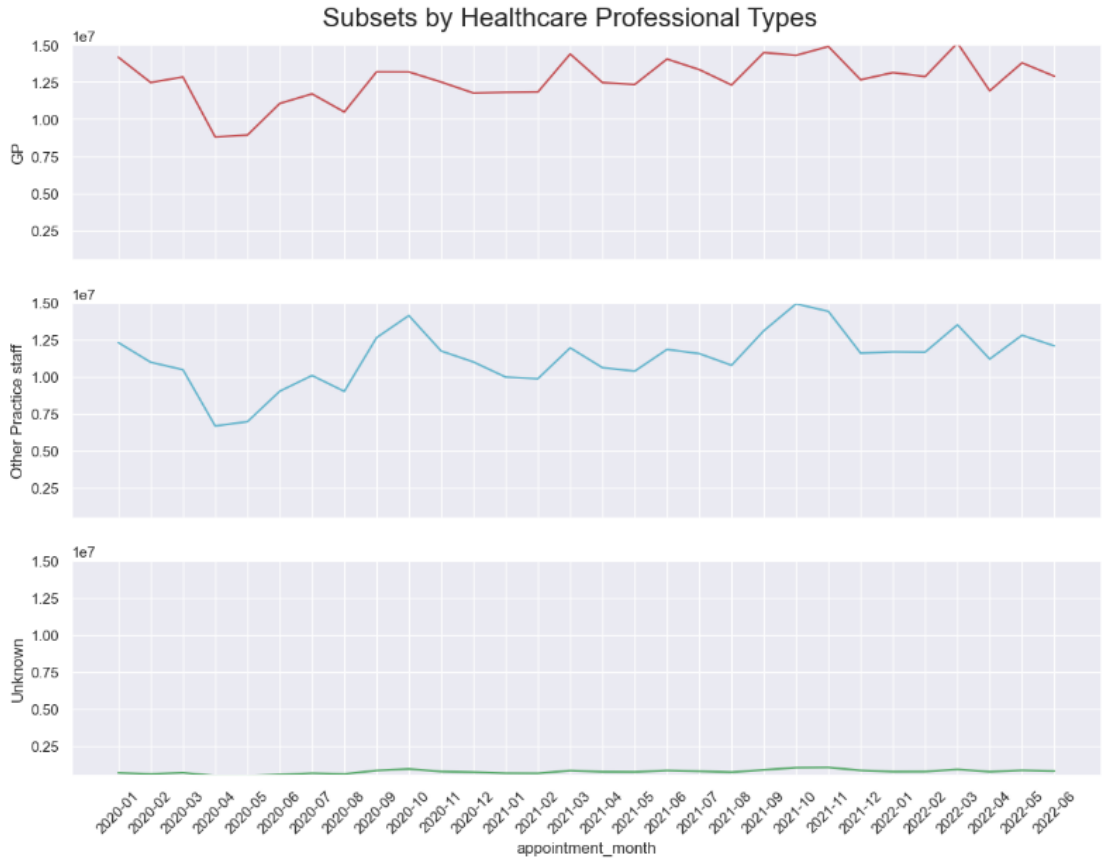
As we could say; The NHS needs to confirm the for accommodate a maximum number such as 1,200,000 appointments per day at this stage. However, the max utilization checked in the dataset was around 1,013,00



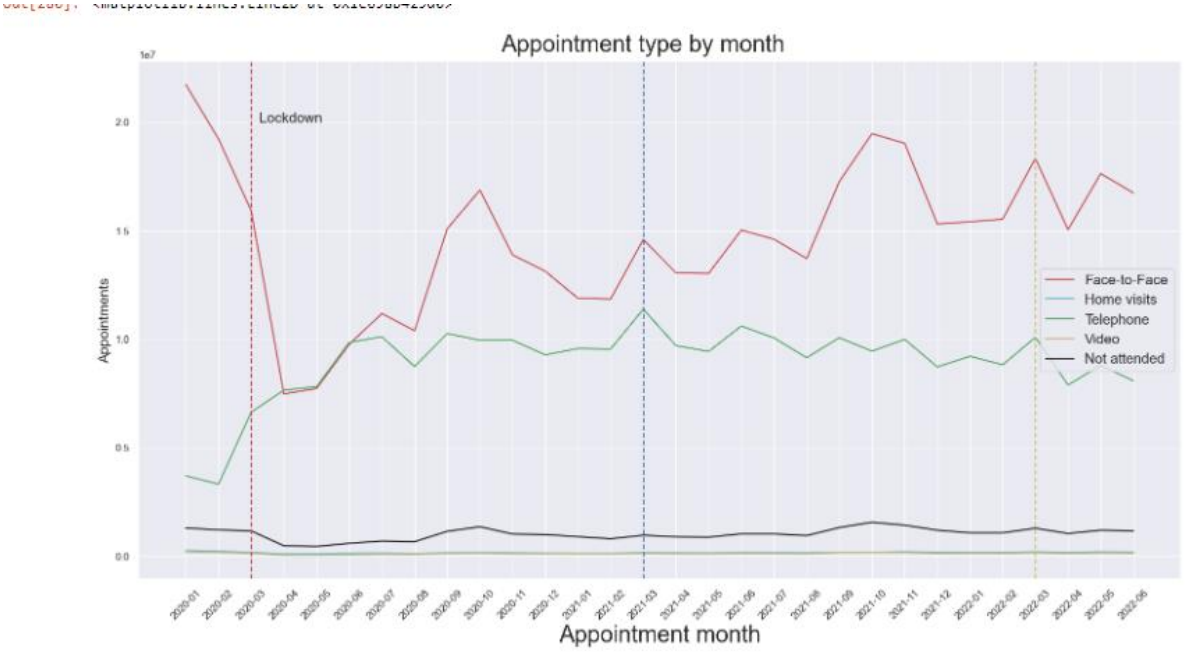
In another way, in April 2020, right after Covid Lockdown. The NHS figures reached the maximum number then it followed the usual up and downs during the dataset times. To enable a better understanding, capacity limits could be calculated by practice, knowing staff availability/demand /budget etc.



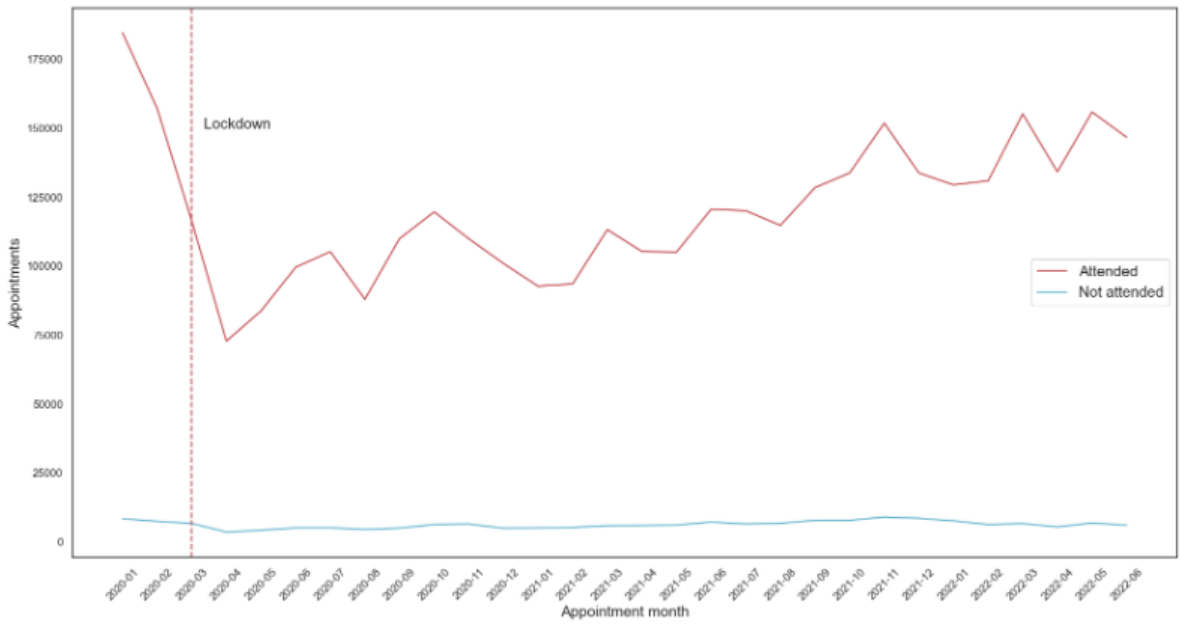
- How do the healthcare professional types differ over time?



Patterns follow among the GP and OPS, which I believe it is related to the demand except for 2020-09. Healthcare professionals seem to be constant over time, there is no a big variance.

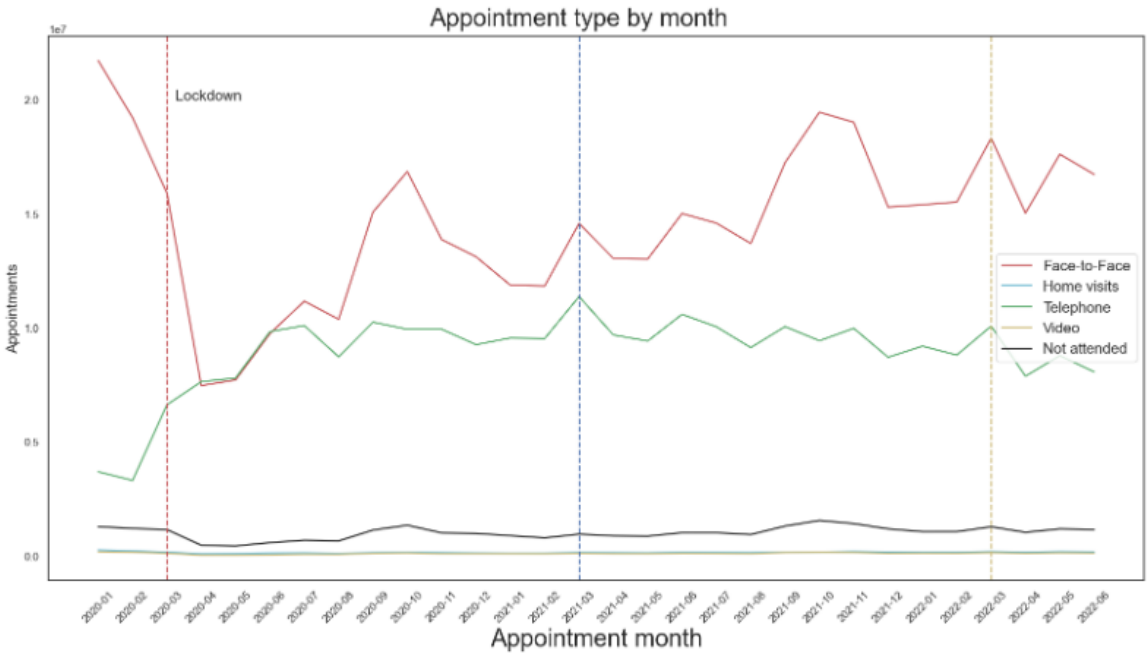


- • Are there significant changes in whether or not visits are attended?



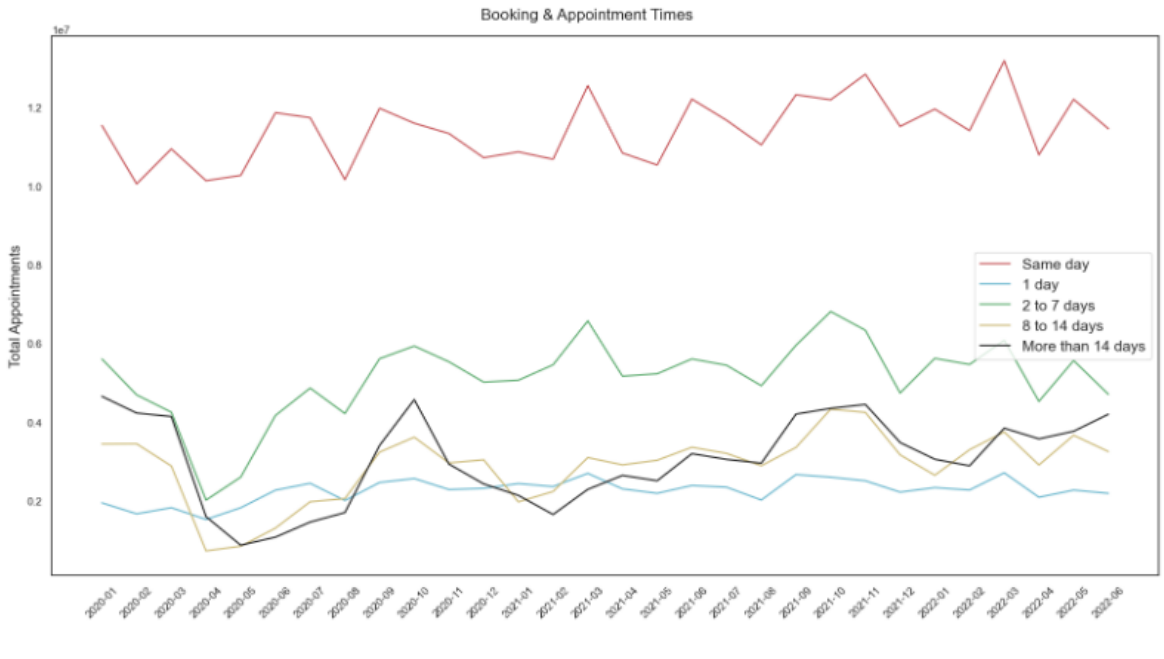
According to the fig, there is an increase for the attended category.

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



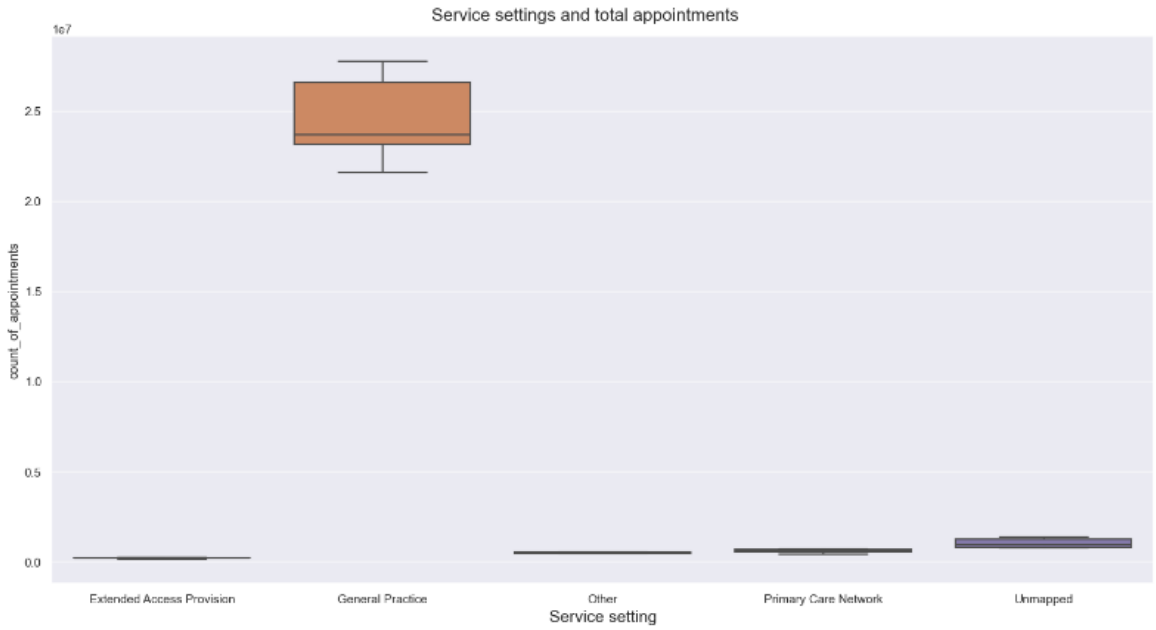
After the lockdown; the face-to-face category is falling down to the pandemic and late-summer 2020, starting to increase until 2020-09.I believe the updated covid regulation on this data affects the face-to-face until 2020-12. On the anniversary of Covid, all appointment types are following the same pattern until 2021-07 and face to face category is reaching the peak.

- Are there any trends in time between booking an appointment?

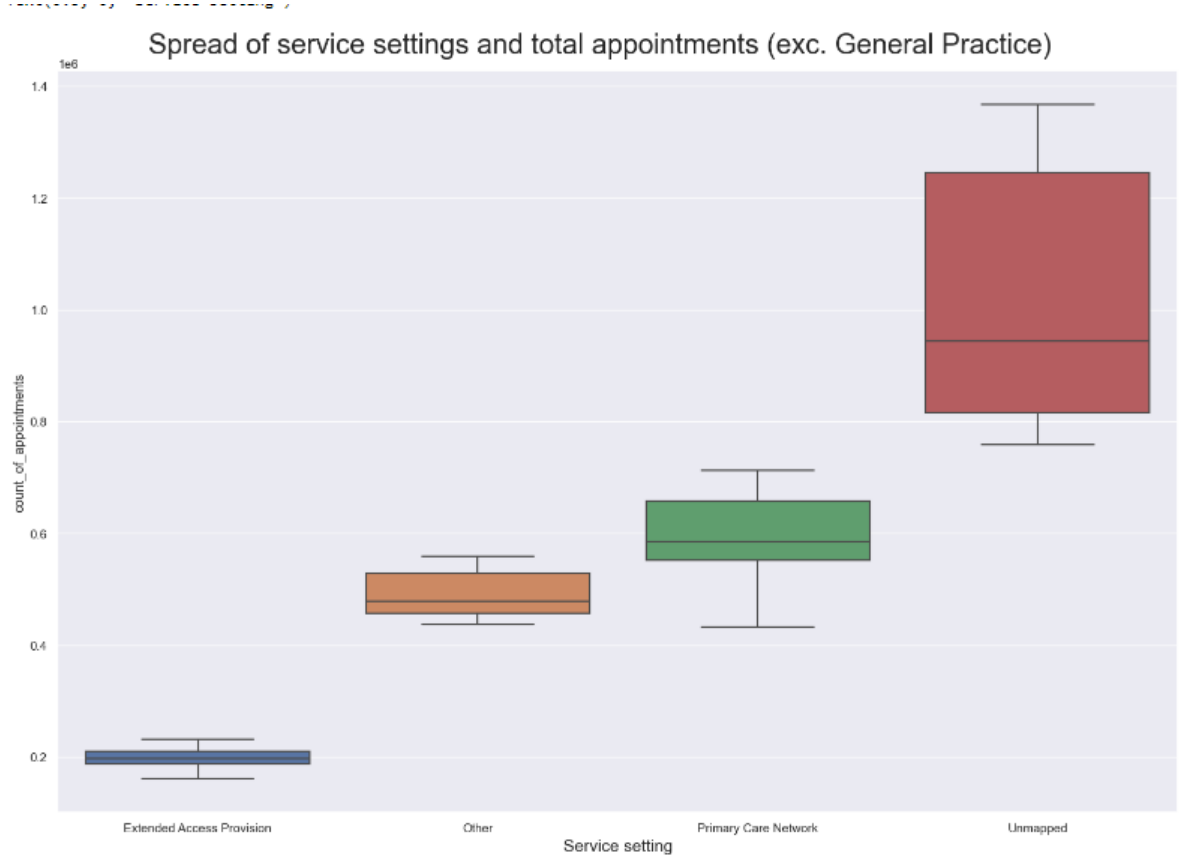


Same-day appointment bookings are the first in terms of volume and similar patterns follow the rest of the chart except 2020-04.

- How does the spread of service settings compare?
  GP is the top in this category and chart in terms of volume.



To find out the rest of the service setting categories, there is another fig below which was prepared without GP.



Unmapped has the second list and PCN is following the list with the third order.

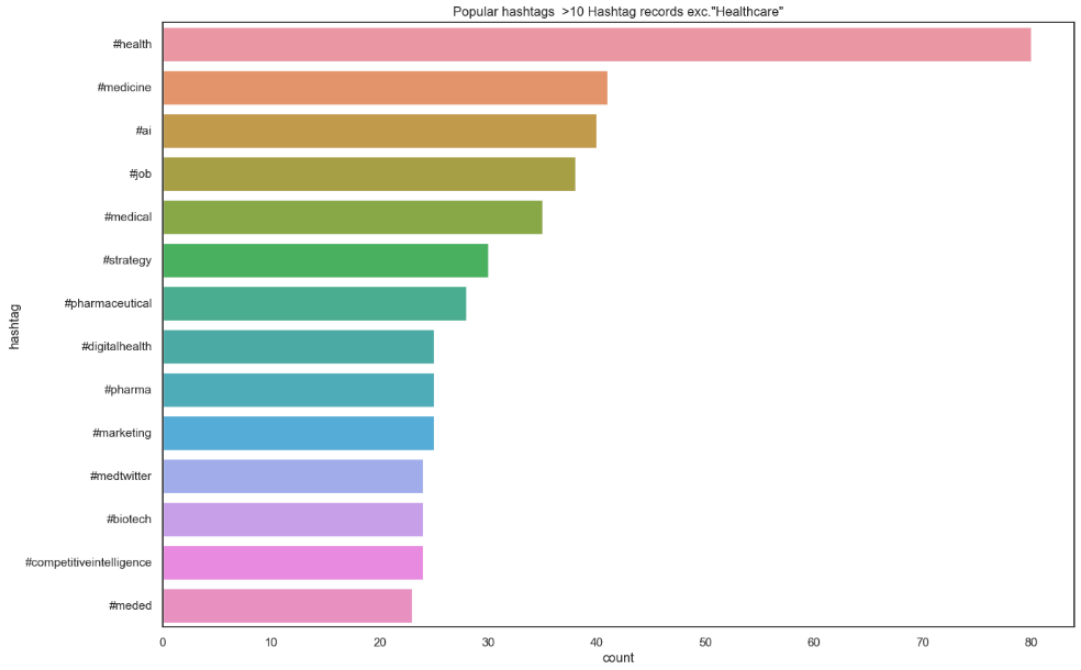
3.2 Social Media Reflection/Twitter Analysis

What are the top trending hashtags related to the HealthCare in the UK?

```
Out[99]: #healthcare                    716
         #health                        80
         #medicine                      41
         #ai                            40
         #job                           38
         #medical                       35
         #strategy                      30
         #pharmaceutical                28
         #digitalhealth                 25
         #pharma                        25
         #marketing                     25
         #medtwitter                    24
         #biotech                       24
         #competitiveintelligence       24
         #meded                         23
         #vaccine                       18
         #hiring                        18
         #news                          17
         #machinelearning               17
         #technology                    17
         #coronavirus                   16
         #womeninmedicine               16
         #covid                         16
         #competitivemarketing          16
         #wellness                      15
         #healthtech                    15
         #doctorofveterinarymedicine    14
         #science                       14
         #medicare                      14
         #covid19                       14
         dtype: int64
```

Healthcare is the top trending topic in terms of hashtags and the below chart indicates the rest excluding " #healthcare":

```
Out[106]: Text(0.5, 1.0, 'Popular hashtags  >10 Hashtag records exc."Healthcare"')
```



3. Further Recommendations

- Currently, the NHS has the right staff levels and enough to cover all appointments recorded over the data set provided the NHS can accommodate a maximum of 1,200,000 appointments per day and the max utilization checked on the data series 1,013,502.3 due to the fact that the increase isn't recommended unless the NHS want to increase the limited appointments.
- GP is the service setting most attended and in face appointments followed by telephone is the preferred appointments.
- Overall peaks October 2021 and March 2022 due to the Covid process (the second lockdown was announced in October 2021 and in March 2022 the third lockdown restrictions started to be lifted.)
- Except these peaks the number of non-attended appointments seems to stay constant over time.
- Same-day appointments closely followed by 2 to 7 days are the most common time between booking and appointment.

Recommendations for further exploration:

- Analysing more data outside the Covid 19 time period for better understanding.
- Limited conclusions can be drawn regarding the utilization of resources.
- Knowing that appointment numbers peaked on Mondays and drop at weekends suggests that more weekend resources could be a solution.
- Appointments being missed is on the gradual increase.