

# **Assignment: Diagnostic Analysis using Python**

**Omer Khan**

**Word count: 890**

## **Overview:**

The national Health Service (NHS), a publicly funded healthcare system in England, incurs significant, potentially avoidable, costs each year when patients miss General Practitioner (GP) appointments.

As part of a data analyst team, this report investigates into the possible reasons why appointments are missed, whilst trying to understand the proper utilisation of resources and adequate capacity in the networks.

## **Approach:**

The following three publicly available datasets were analysed:

- Actual duration – details of appointments made by patients including regional information, date, duration and number of appointments pertaining to a certain class.
- Appointments regional – details on the type of appointments made by patients, including regional information, the month of appointment, appointment status, healthcare professional, appointment mode, the time between booking and the appointment, as well as the number of appointments pertaining to a certain class.
- National categories – details of the national categories of appointments made by patients, including the regional information, date of appointment, service setting, type of context, national category, and the number of appointments pertaining to a certain class.
- Tweets – Data related to healthcare in the UK scraped from Twitter.

A Git repository has been created for this project where all the relevant files including the above mentioned datasets and the notebook containing the Python code have been uploaded, in order to ease collaboration with the data analyst team. Git hub allowed each team member to work together controlling changes over time and learning from and with others facilitating a thorough analysis.

Pandas, Numpy, Seaborn and Matplotlib were used as part of the Python analysis.

## **Insights:**

Initial checks were performed to view the data and it was found that the data included 18 national categories, 106 locations, 5 service settings and 3 context types.

NHS North West London ICB was found to be the location with the highest records. The most common setting was GP setting and the highest context was Care Related Encounter.

The data was wrangled by applying filters to focus only at data from 01.08.2021 to 30.06.2022 to look at seasonal trends. It was discovered that November and October 2021 were the two busiest months respectively.

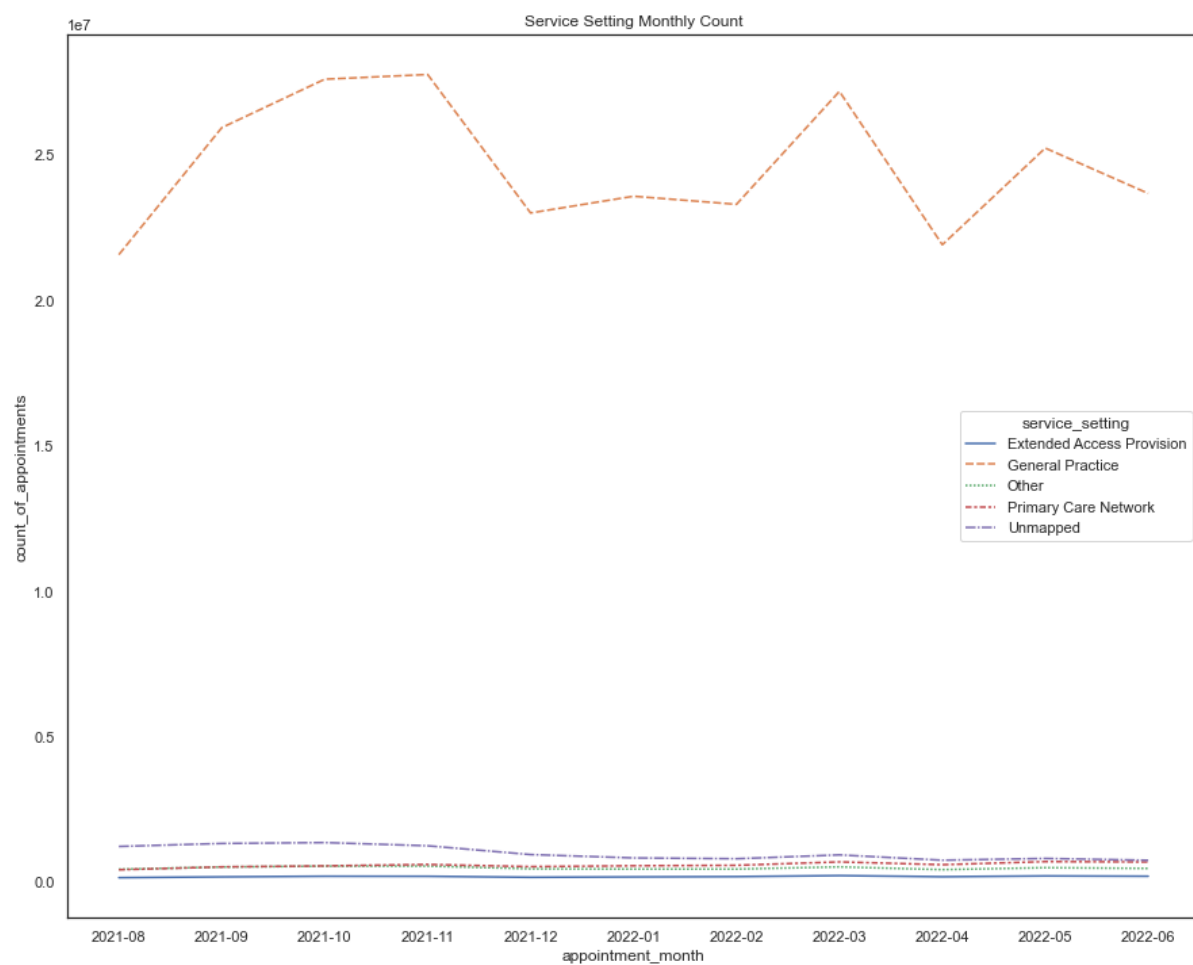


Fig 1. Service setting monthly count

Figure 1 above highlights the trend that GP services are by far the most that patients resort to, as well as indicating the busier periods to be around October and November, with December 2021, April 2022 and August 2021 being the least busiest.

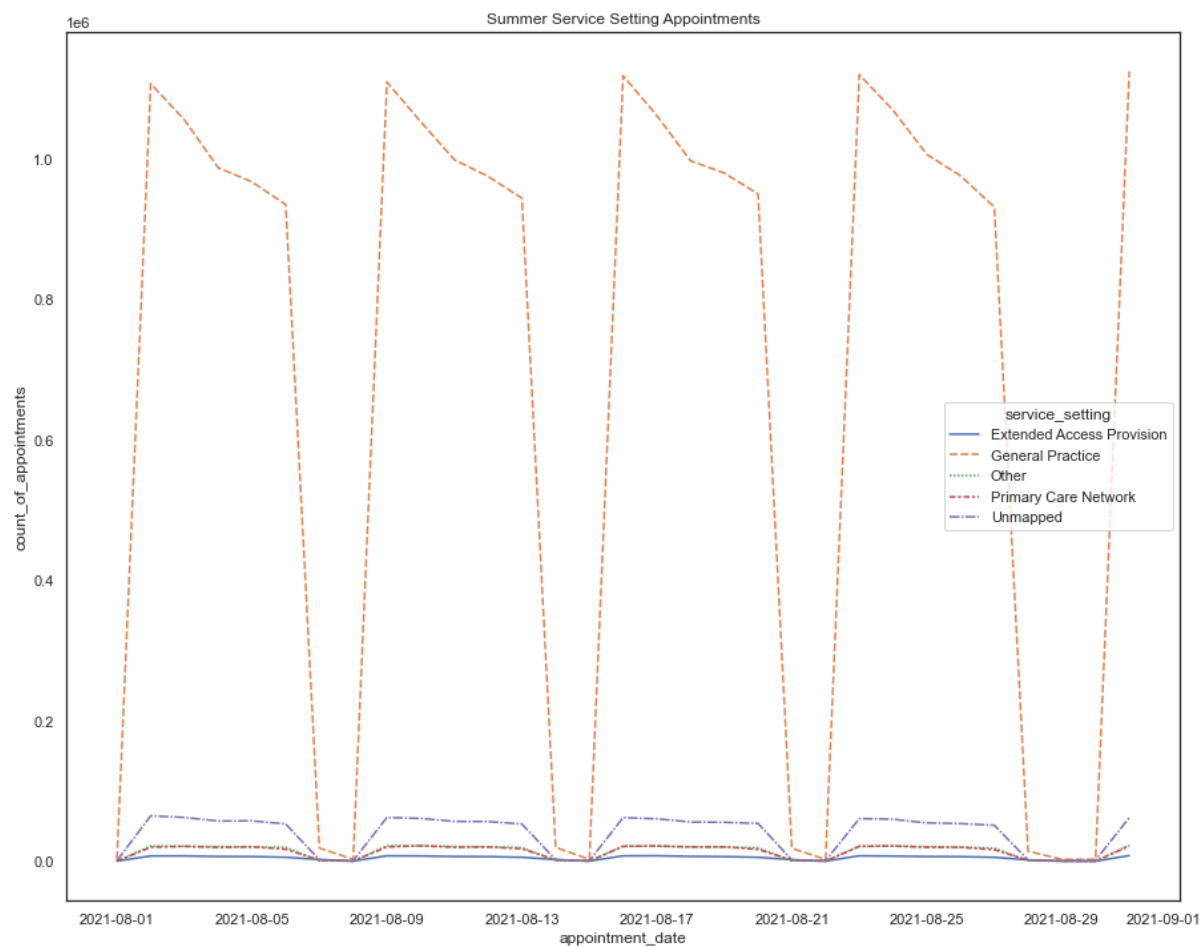


Fig 2. Summer Service Setting Appointments

Looking at the seasonal trends in more detail, as in Fig 2, shows that there is a consistent pattern throughout the year for Mondays to be the busiest day for GP services.

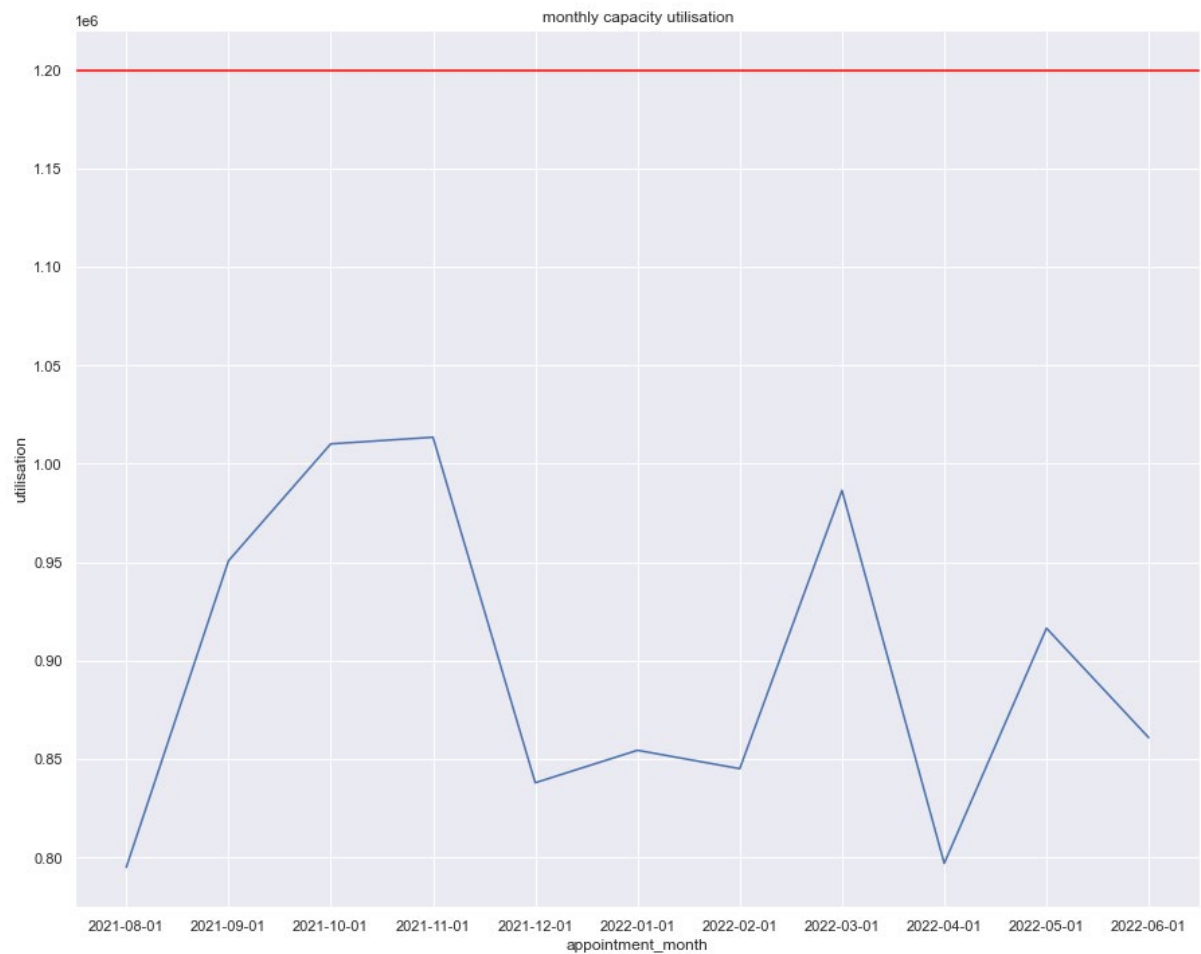


Fig 3. Monthly Utilisation

To understand the utilisation of the services, Fig 3 was plotted. The red line at 1,200,000 shows the daily NHS appointments availability, which was no way near utilised to its full extent. This should be further explored to better understand the reason for this. There is also a need to understand the 1,200,000 daily limit that NHS claims – is this based on current staff capacity and resources?



Fig 5. Missed appointment by HCP Type

Fig 5 was plotted to understand which settings had the most missed appointments. As Fig 5 suggests this was mostly in Other Practice staff with a peak seen in October 2021. This could perhaps have been due to the reduced number of services being offered at that time due to the ongoing Pandemic. Data should be looked at for previous years to determine if this was affected by Covid-19 or not.

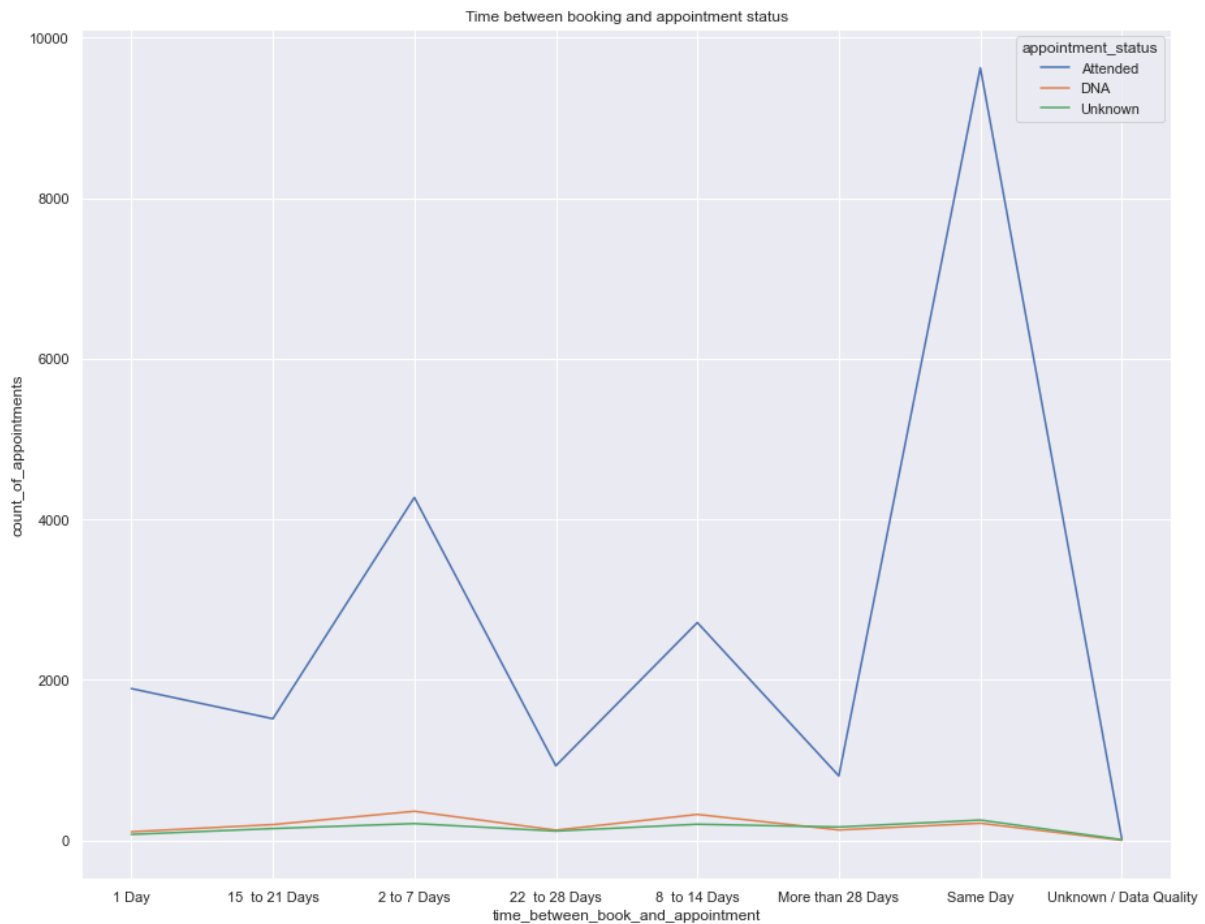


Fig 6. Time between booking and appointment status.

Further evaluating missed appointments, Fig 6 was plotted to observe possible relation between the duration between booking the appointment and attendance. It can be clearly seen from the above graph that most of the appointments made were on the same day and had a high rate of attendance. The orange line indicates that the highest number of missed appointments were those booked between 2-7 days and 8-14 days. The reason for this should be analysed further – How were patients informed of the appointment and did that factor into this?

It was also observed that most appointments that were missed were face to face appointments – this could be valuable as the attendance for other appointment types was high. Since this data has been affected by the Pandemic it would be worth comparing it with previous years to see if face to face appointments were missed due to Covid guidelines or other factors.

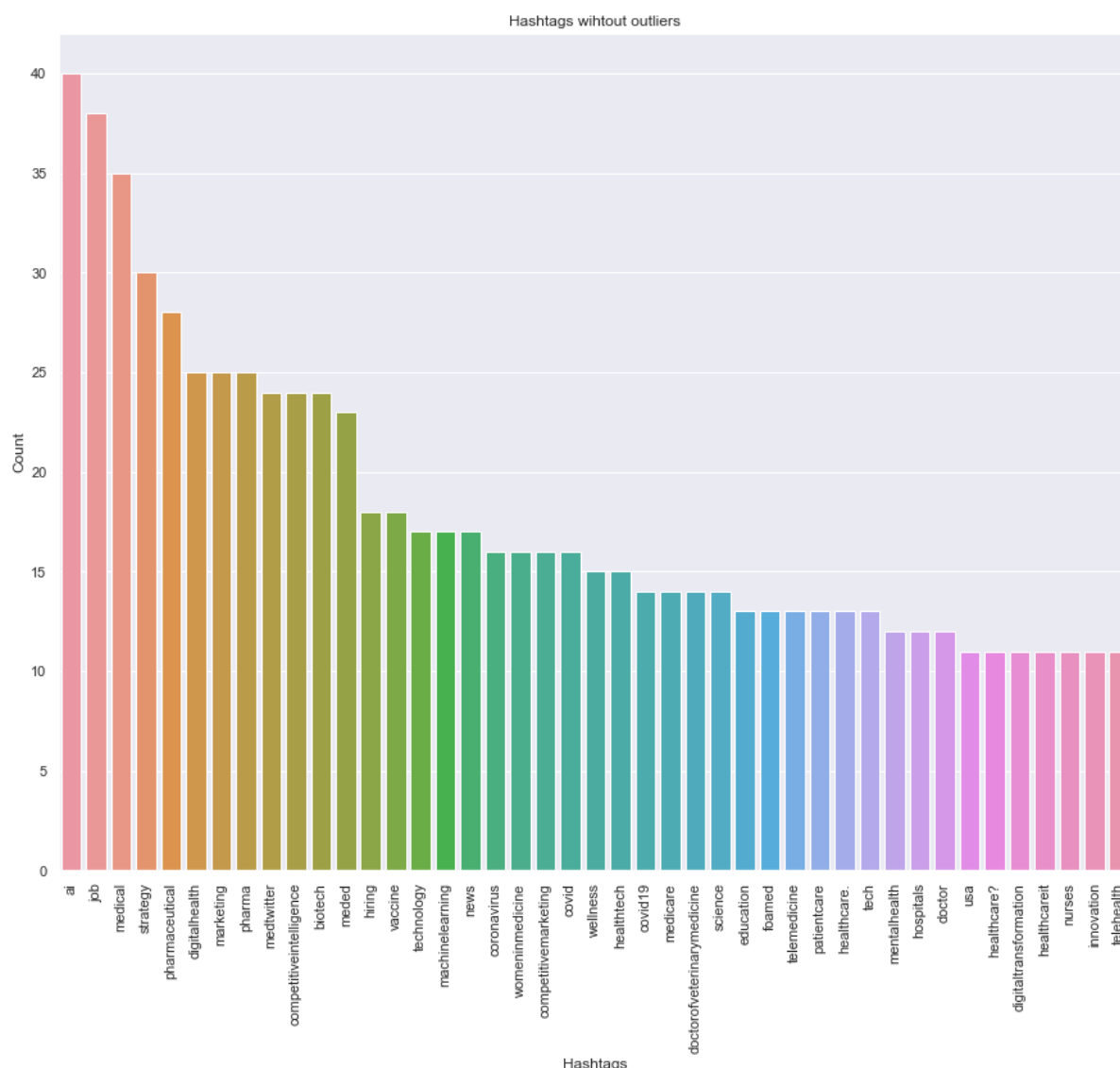


Fig 7. Twitter Hashtags without outliers

A twitter hashtag analysis was also conducted during this time period to spot any trends related to the NHS. Initially it was observed that the top 3 hashtags were #healthcare, #health and #medicine. However, constructing a boxplot suggested these to be as outliers and removing them resulted in Fig7 (above). Nonetheless, understanding the nature of social media and Twitter, these hashtags should not necessarily be treated as outliers.

### **Key highlights and suggestions:**

This analysis has brought in light areas for the NHS to look at:

- Detailed analysis in to why daily appointments are far below the claimed capacity of 1,200,000.



- Further evaluating the reason for missed Face to face appointments and those appointments which are booked 2-7 and 8-14 days in advance.
- Observe twitter trends and analytics to better understand the patients and to track NHS's performance and progress.