NHS Python Data Report

By Geoff Thomas Widdowson

Beng(hons) IPS

March 2025

Introduction/Problem Statement

In this assignment I will look at a number of huge NHS databases to understand trends in the data over a period of time using Python. I will look at the data with regards to appointments in GP practices to better understand what the situation is.

Through my analysis I will seek to investigate two main questions being asked by the NHS:

1. Has there been adequate staff and capacity in the networks?

2. What was the actual utilization of resources?

The overall aim is to see if there is a way we can improve things moving forward while gaining insight into the workings of the GP practices in England.

Analytical Approach

In looking at the problem presented my approach was to initially import, consider and begin exploring all the available data using Python. The Meta data stated the sets had been cleaned so I did not spend long looking much further at cleaning as there were 4 extensive data sets provided. I did find and drop some duplicates in the data as I began exploring.

In looking at the sets of data I discovered they contained various key pieces of information over different time frames as follows:

1. 'Actual Duration' (AD)
   - 168 million total appointments (rounded)
   - Dec 2021 – June 2022
   - Key Data:
     - 106 unique locations
     - Appointment duration

2. 'National Catagories' (NC)
   - 296 million total appointments (rounded)
   - August 2021 – June 2022
   - Key data:
     - 17 types of appointment
     - 5 service settings
     - 3 context types
     - 106 unique locations

3. 'Appointments Regional' (AR)
   - 743 million total appointments (rounded)
   - Jan 2020 to June 2022
   - Key data:
     - 3 discrete health care providers
     - 3 appointment status points
     - 5 types of appointment
     - 8 waiting times on appointments

4. 'Twitter' Data relating to tweets and social media insight.

After initially looking though the data to understand all these key areas I began to gain insight relating to the main three data sets with appointment counts.

I decided to focus on the AR data set as it gave a better picture over a wider time frame.

While the AD and NC data overlapped in time and could potentially be merged around the Sub ICB grouping - I did not do this as I was concerned the data quality could skew and invalidate this analysis. Almost 24% was grouped into the Unknown/ Data Quality Issue which when you consider the potential time allocation of over and hour per appointment is a serious issue in terms of resource allocation!
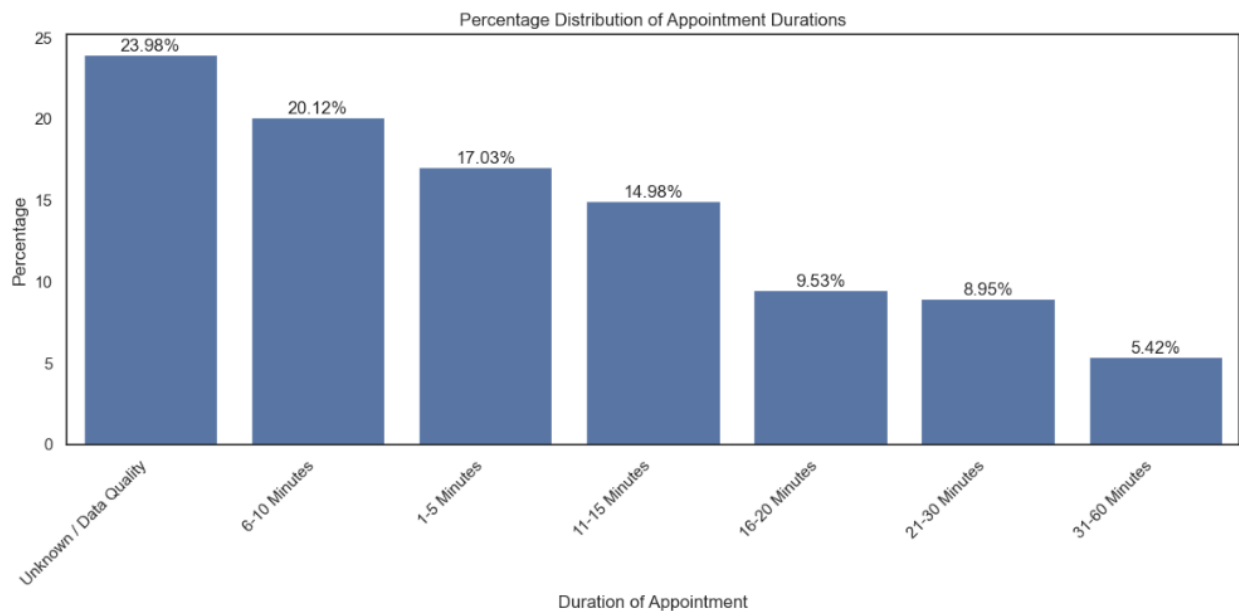
No data was supplied in relation to staffing levels so I had to use the number of appointments as an indirect proxy for staffing levels.

I also considered looking further at wait times between appointments and appointment duration and bringing in external geographical data to help with this but in the end ran out of time to investigate further.

In focusing on the AR data I took a national approach and did not separate by ICB as I did not have any information relating to what this and while location is obviously extremely important I did not have time to find further data on ICB codes and aggregate the date further to draw a more granular approach geographically.

I used a variety of code in python using things learnt on the course with help from tutors and using claude/ anaconda assistant to write code and make further investigations to visualize my findings more pleasingly.

Investigation:

Initially I looked at the overall appointments in the National Catagories set. Here I found that there was an interesting pattern over time concerning the data on a weekly basis.
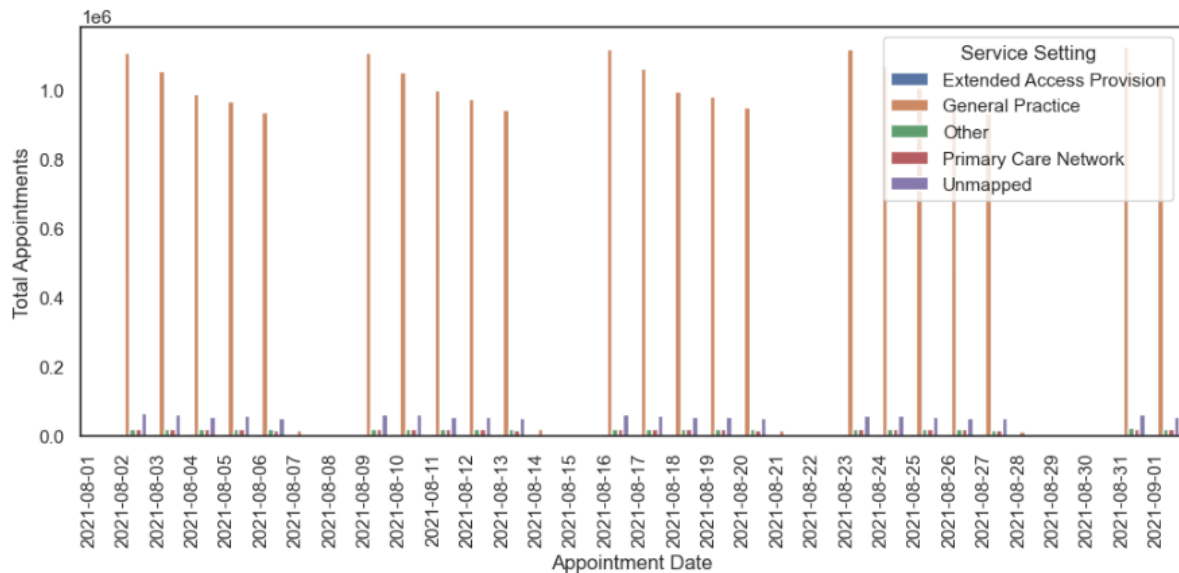
We can clearly see not only more appointments on Mondays but also no appointments over bank holiday weekends. This trend continued throughout the data set and was considered more accurately with the line graph below. We can also see there are the markings of changes as we move through time in terms of the peak total appointments per service setting.
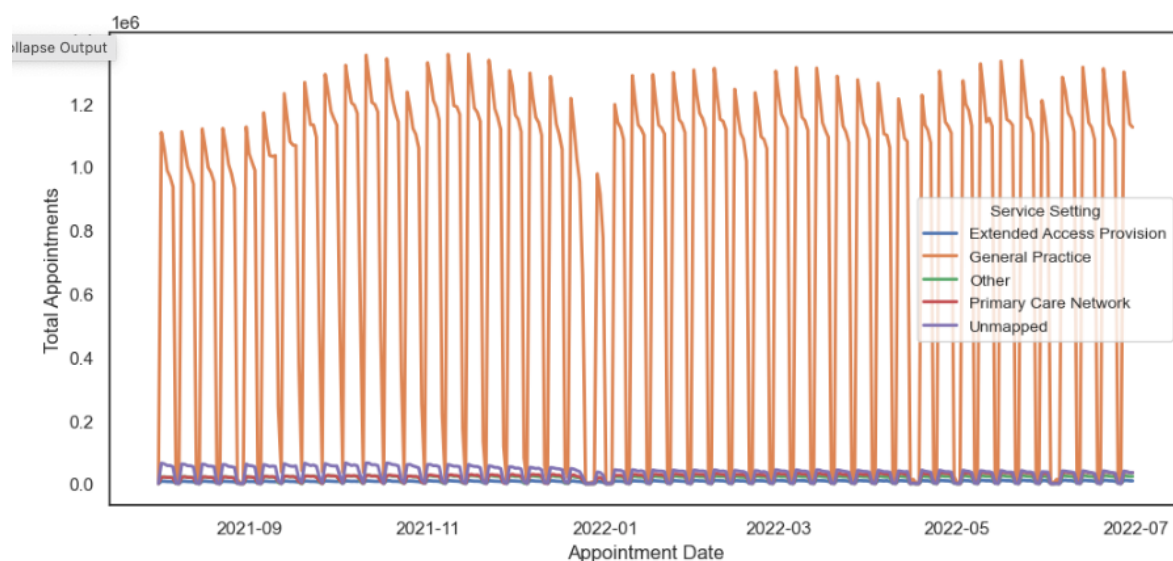
Seasonal analysis of the National Categories data set and working out the average daily NHS appointments per season based upon a 5 day working week led to discovering the increase in average daily appointments from summer through to Spring.


NC Seasonal Analysis 2021/22
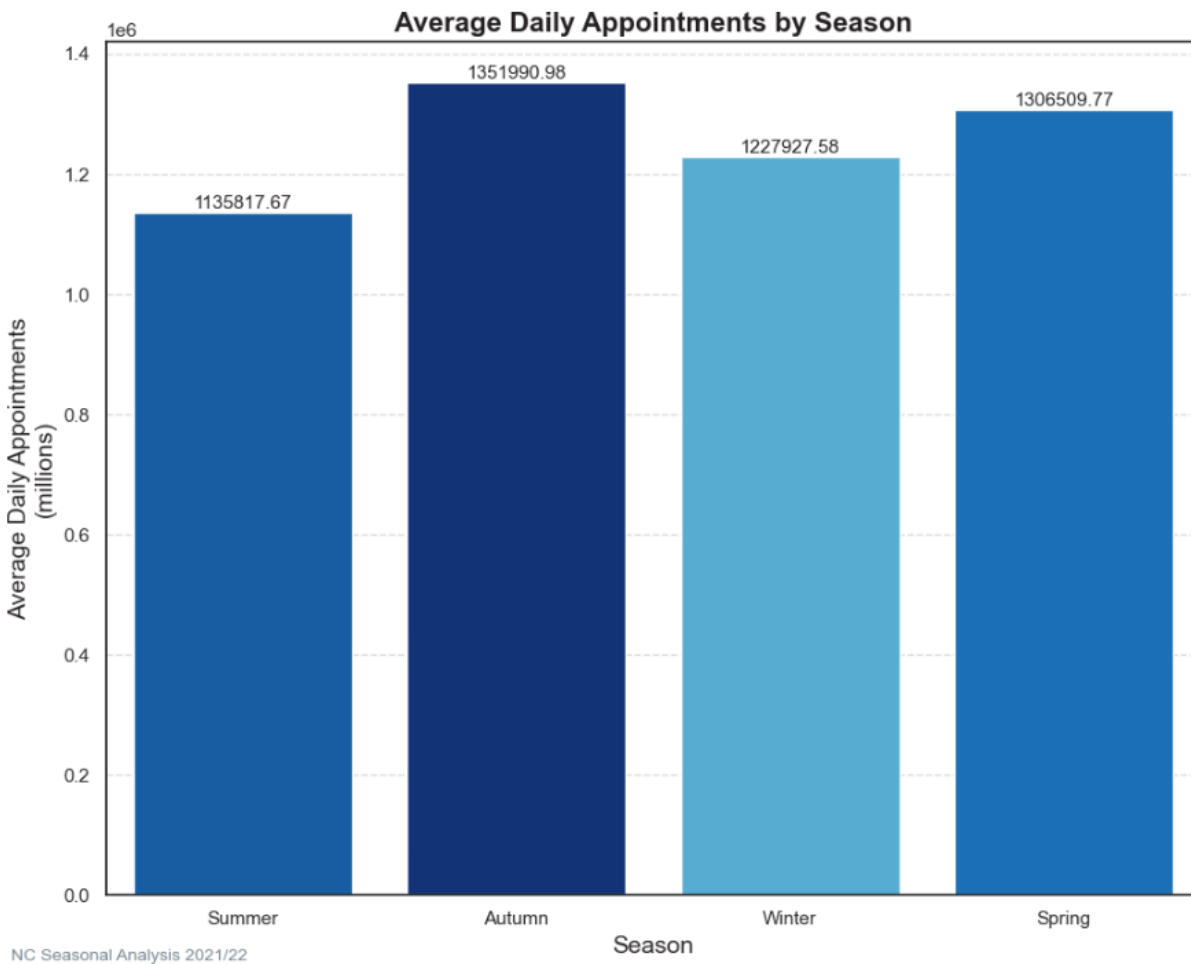
While considering this I wanted to look further at a broader range of time so I considered the AR data set in more detail next.

I started by looking at appointment over time with regards to the HCP service data.

Total Monthly Appointments by Service

Jan 2020 - June 2022 (AR Dataset)
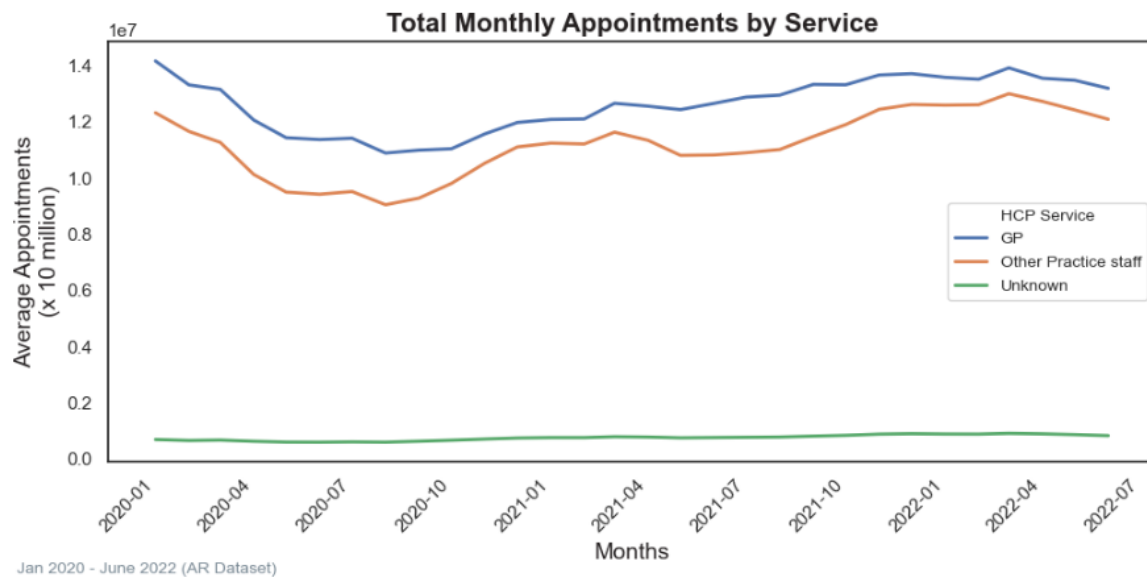
We can see an increase in appointments over time with practice staff accounting for almost the same appointments as GPs!

Going back to NC data I looked further into national categories visualizing the top 5 categories over time.

**Distribution of National Categories Over Time**
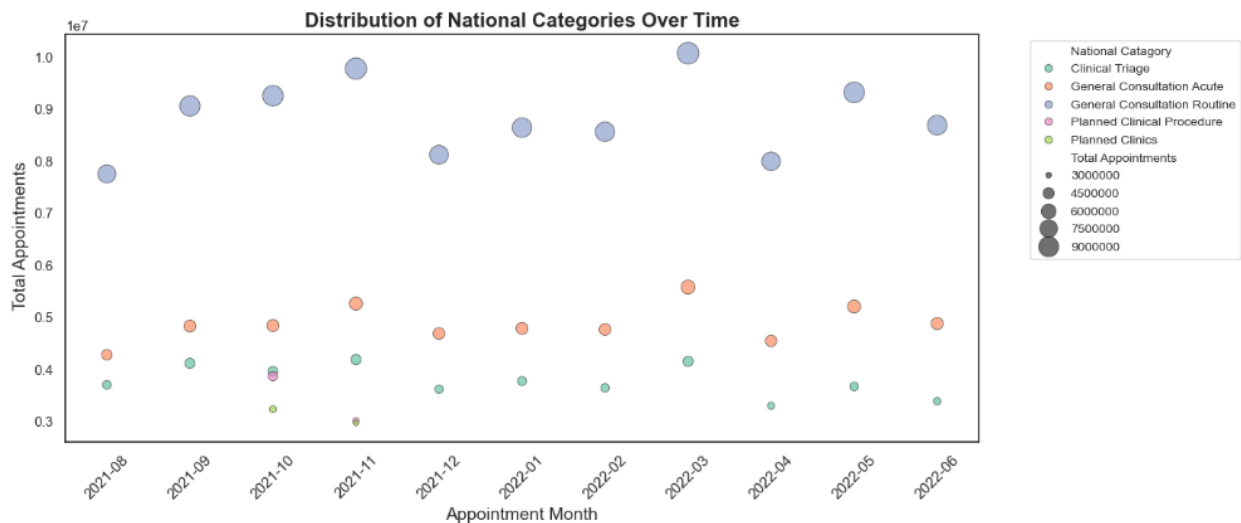
Looking at the distribution of total appointments across the national categories and we can see it following a similar pattern.
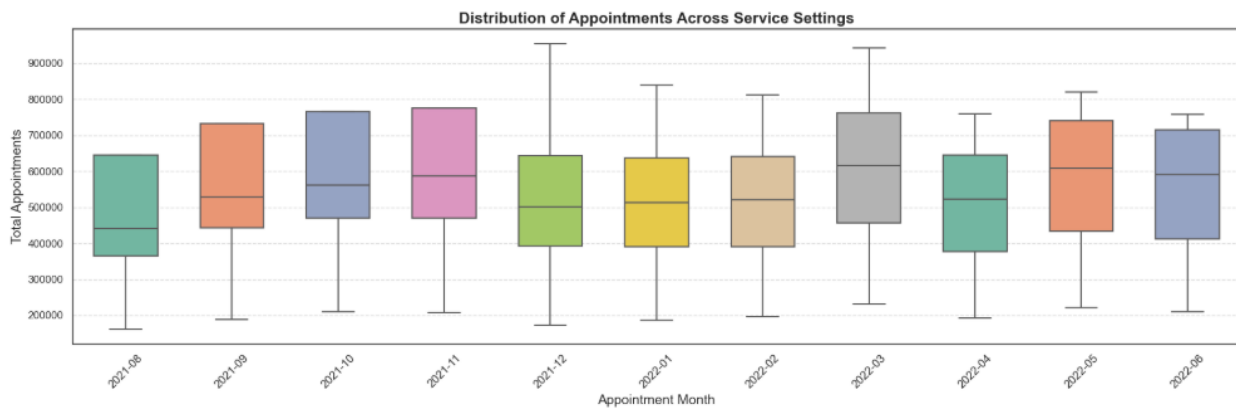

We can see the NC data set gives us a broader picture in terms of catagories relating to all the other staff working as part of practices.

In going back to AR I considered the HCP Service with regards to the % change and we can see wild swings over time showing - the system is under pressure!?
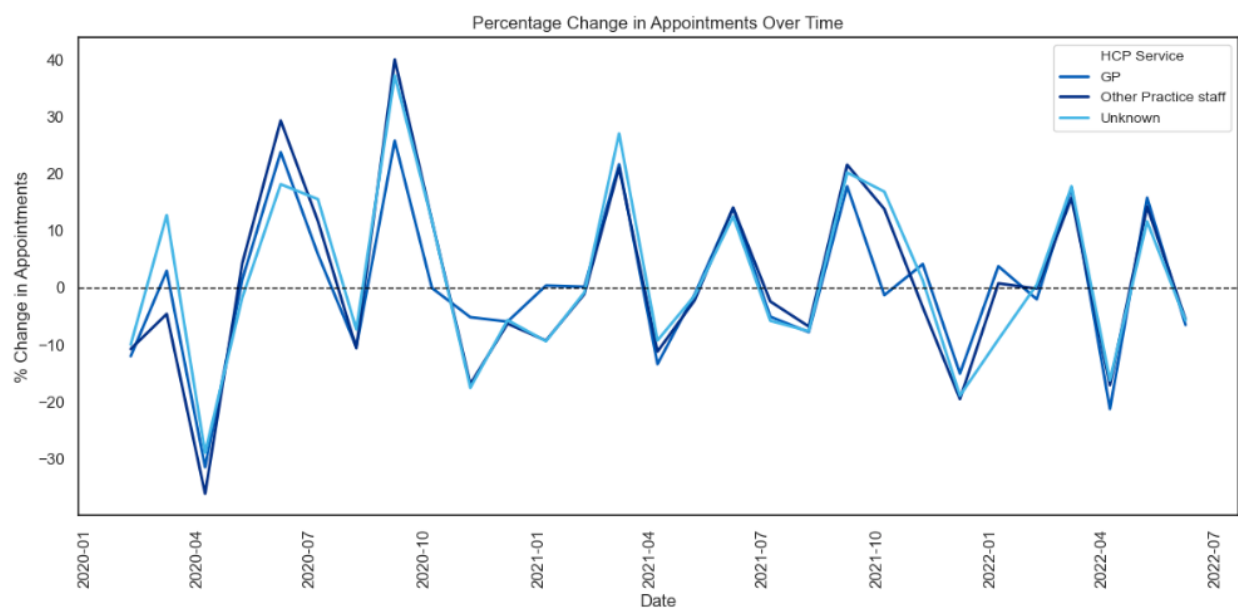
Continuing to work while considering actual working days at practices (5 day working week, plus Bank holidays and base Annual leave of 27 days) I looked further at the total appointments over time.
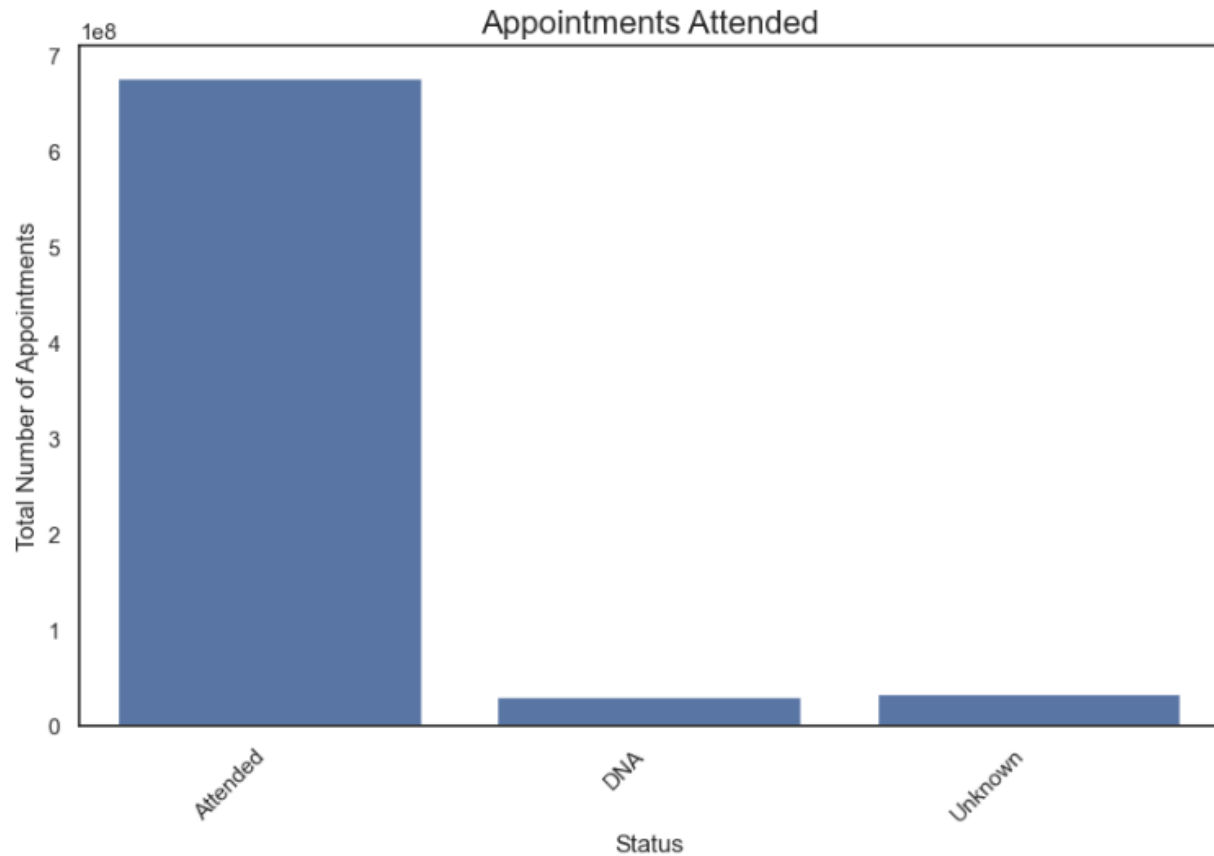
We can show there is a clear trend here in terms of and increase in appointments over time.
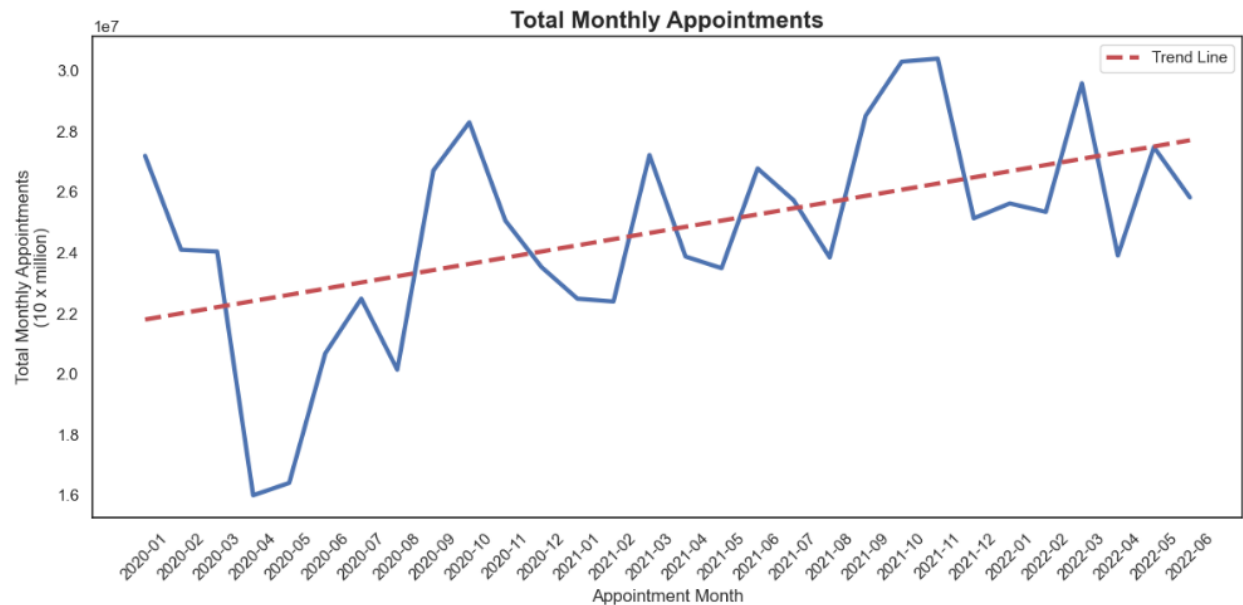

Looking at the monthly utilization of the services it also becomes clear that the GP practices are going well above their daily appointment target of 1.2 million despite this!
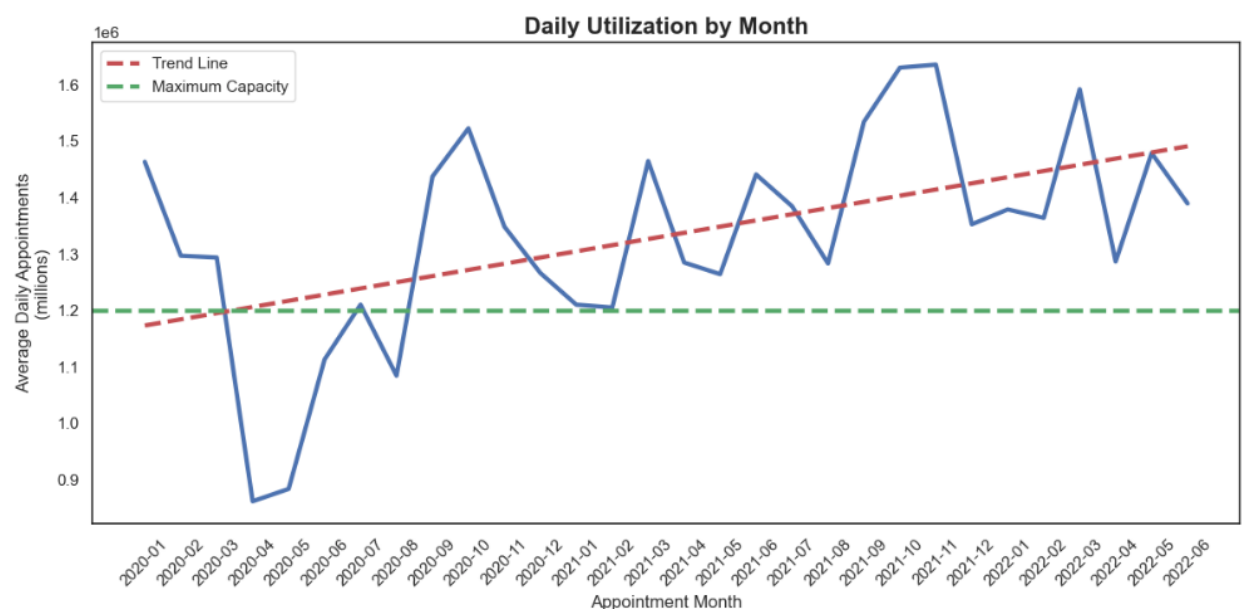
Good news? We can see the majority of appointments are being attended.

Waiting times are good with around 50% of the total patient appointments happening the same day the are booked with less around less than 25% taking over a week.
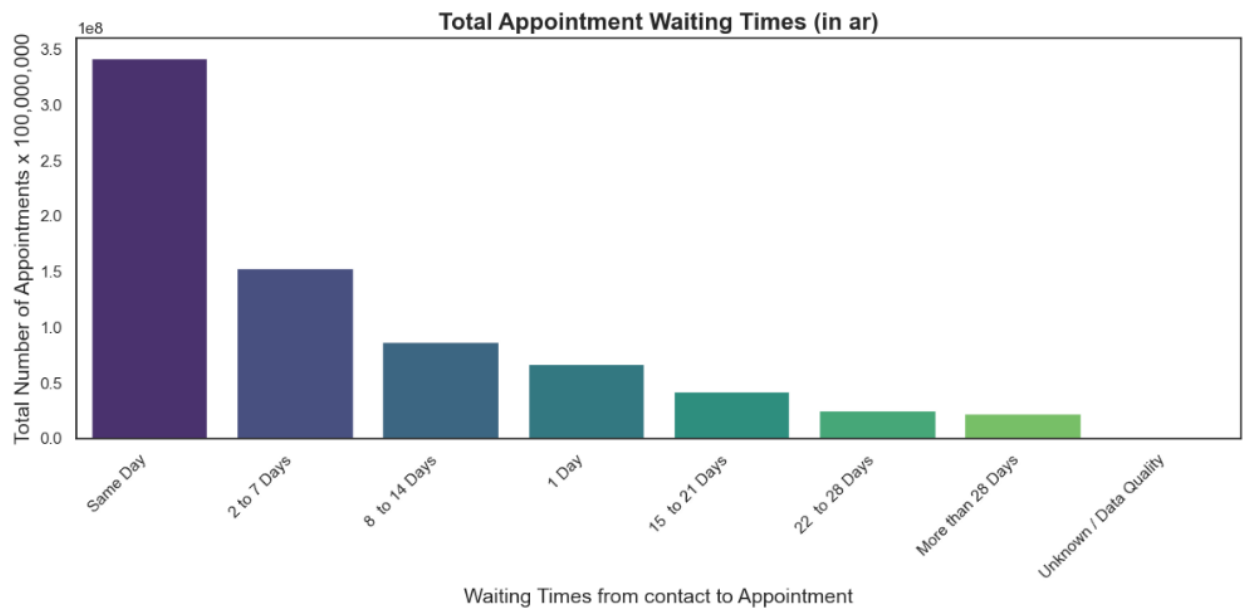
Looking further I found that there has been a significant increase in Telephone appointments over time and while this was probably driven by the covid 19 pandemic it could be something of significant benefit in moving forward.

Appointment Trends by Type
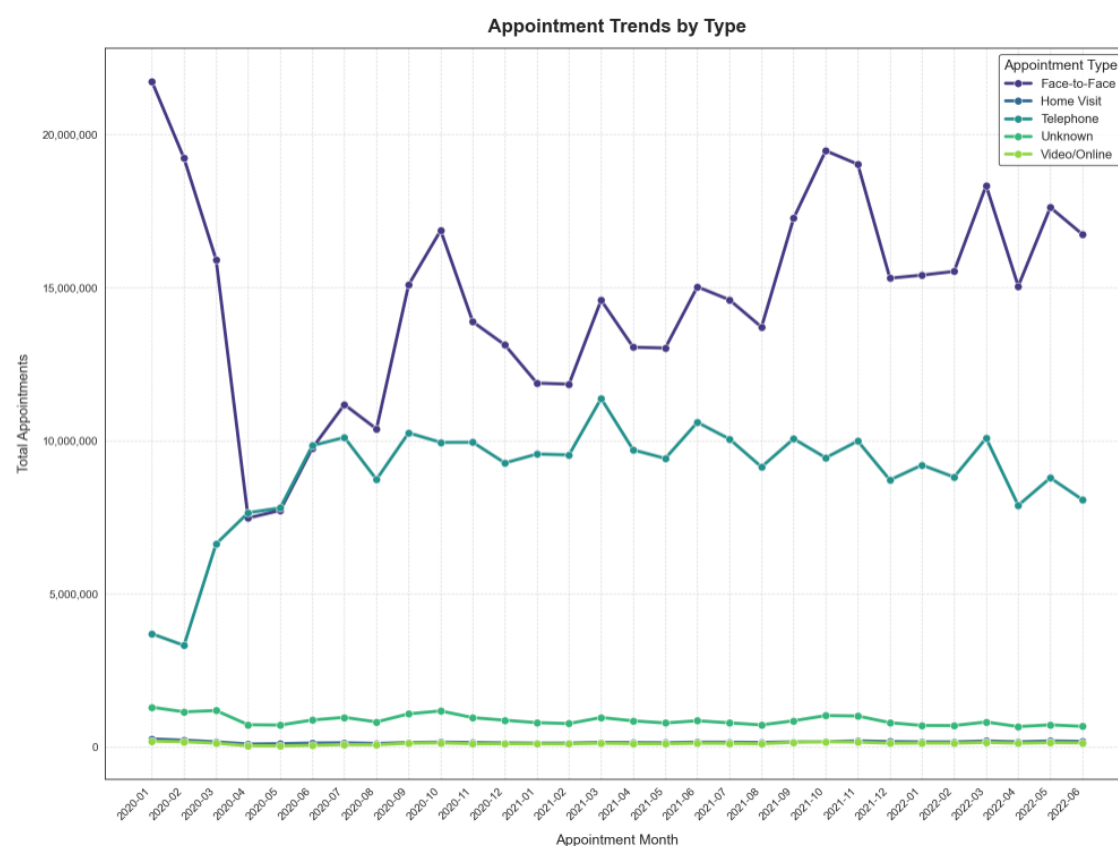
If I had more time I would investigate this further to understand trends around other variables such as waiting time, appointments attended, locations and national categories.
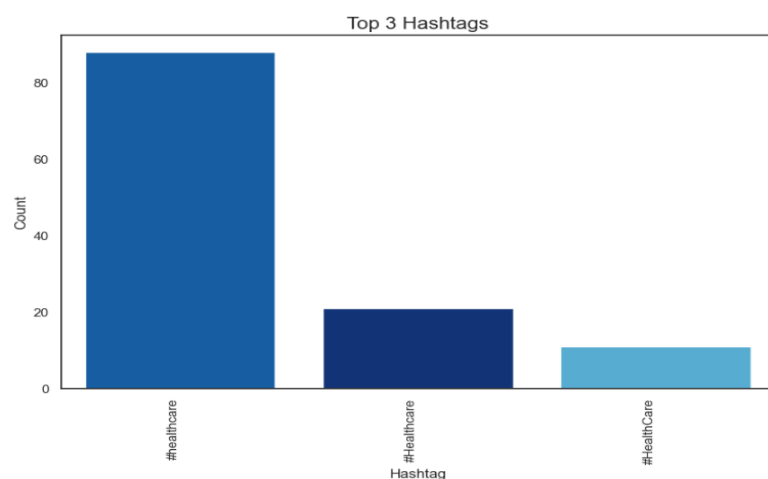
Top 3 Hashtags

I also considered the twitter data set but there was not a great deal to work on or enough time to dig deeper I did how ever discover the top 3 hashtags!

#healthcare

The future.

It is clear there is not adequate staffing within the GP practices and it is not yet completely clear how we can utilise existing resources further given that there are inconsistencies around some of the data that has been collected.

Telephone appointments are a good utilisation of resources that may be able to help take some pressure off the system moving forward with further investigation.

Recommendations:

Look at how existing staff can be distributed over the week to handle increases in weekly demand

Consider further appointments on the weekends to ease weekday demand

Investigate further Telephone appointment utilisation

Gather more granular data for HCP's

Look to gain location based data insights