

Introduction

The National Health Service (NHS) is the publicly subsidised medical and healthcare service within the UK. The NHS is under intense pressure to discover better and more effective ways to provide care, reduce wait times, and focus health services and preventative efforts on the populations that need them the most.

Specifically, specific trends or patterns could be highlighted through data and data-driven systems to explain and understand the reasons for missed appointments. The government can choose the best way to handle this problem.

Analytical approach

Evaluating the data to guarantee quality and validity before any analysis would be good. Examples include fixing errors, ensuring that the data is formatted correctly, determining whether any numbers are missing, and much more.

To load and sense-check the data, I first imported the pandas package. The next three DataFrames need to be created for each of the datasets. I used the `pd.read_csv` and `pd.read_excel` since the datasets are of different file types. To get a general overview of each DataFrames I used the commands, `.head()`, `.info()`, `.columns()`, `.dtypes()`. I could be able to see the various columns, the datatypes and more. With this information, I then choose to look at the options or different categories in some of the columns that may be important. For instance, use the code `.value_counts()` to see the number of locations with the DataFrames. Other methods I used are `.count()`, `print()` and so on.

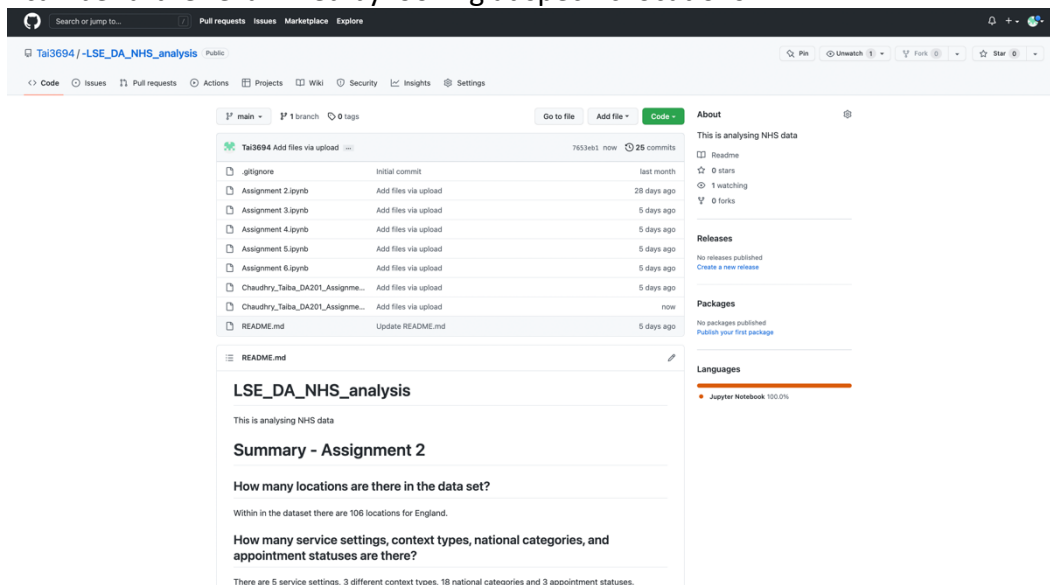
Next, I employed the command `.isna()` to see if there were any missing values. If this was the, I then used `.isna().sum()` to get the total count of missing values within each DataFrame. I found no missing values for all DataFrames and began to look at some descriptive statistics using the `.describe()` command. Specifically, getting a descriptive statistics table for the `count_of_appointments`. There were differences between the three DataFrames, such as the count, mean, and standard deviation. Yet, it could be of interest as we can see that the national categories dataset had the most data available and would be suitable for finding any results or trends.

To start identifying any patterns, I thought it was crucial to work more with the data and find important aspects. When looking at the time of appointments, I decided to change the format of the `appointment_date` to a datetime format so that it is in a more readable string. I used the `pd.to_datetime` command and employed it for all 3 DataFrames. I then sense-checked the DataFrames to confirm whether it was successful. I then used the `.min()`, and `.max()` commands to find the range for appointment dates. I also wanted to see the count of appointments across the different service settings. For this I used the `.groupby()` and `.count()` commands. As a result, I find that general practices are the most popular.

These are just some small ways of manipulating and working with the DataFrames. I used various methods to get an overview of the essential details. Then, being able to adjust the necessary columns or variables to make them more suitable to work with. From here, small

GitHub: [https://github.com/Tai3694/-LSE DA NHS analysis](https://github.com/Tai3694/-LSE_DA_NHS_analysis)

insights can be seen, such as the timing of appointments and the common service settings, which can be further examined by looking at specific locations.



Given this, I choose to set up a GitHub repository to store and update the project files. This is a great platform for archiving, monitoring, and working together on software projects. This would aid in presenting the codes and projects to my team and allows for collaboration.

Visualisations and insights

Creating graphs or other visual representations may be more beneficial when dealing with complex data. This can all be done within Python using the required packages, matplotlib and seaborn. Matplotlib is a Python graphics package for data visualisation and integrates nicely with Numpy and Pandas. Seaborn is another tool for creating statistical visualisations. It incorporates tightly with Pandas data structures and is built upon Matplotlib.

With this, different types of graphs can be created. Yet, I decided to create line plots and bar graphs as they would be the best to make comparisons and look at different variables across time. Given that the main objective is to identify the reasons behind missing appointments, it would be good to see the number of appointments across different variables. For instance, looking at the count of appointments over time for service settings, context type or national categories. To view this, the `.groupby()` and `.sum()` codes are vital as we can see the total count of appointments in the specific grouped variables. To make it easier, I also saved these grouped data as new DataFrames. To then create the line plot, I would specify the necessary information. An example of the code to create a line plot is below.



Figure 1: Line plot for count of appointments for different service setting over time

The x-axis is the period, the y-axis the total count of appointments, the hue specifies the column chosen for colour coding, which in this case is service setting, data is the DataFrame being used, and ci=None is stating for the confidence interval to be 0. Then directly underneath the code, the graph is created. We can also adjust the chart by setting a figure size or colour. With this visualisation, we can see that General practices are prevalent, and we have seen this insight when getting an overview of the data. I then repeated this process to examine national categories (Figure 1) and context type (Figure 2) over time.

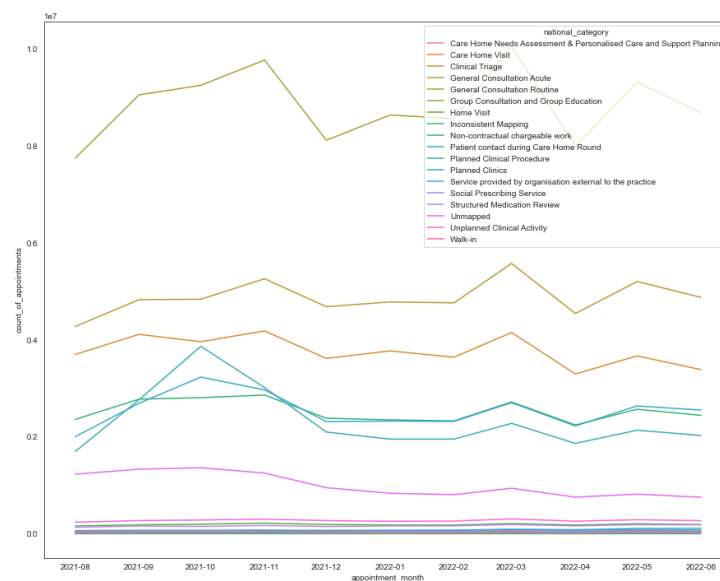


Figure 2: Line plot for count of appointment for different national categories over time

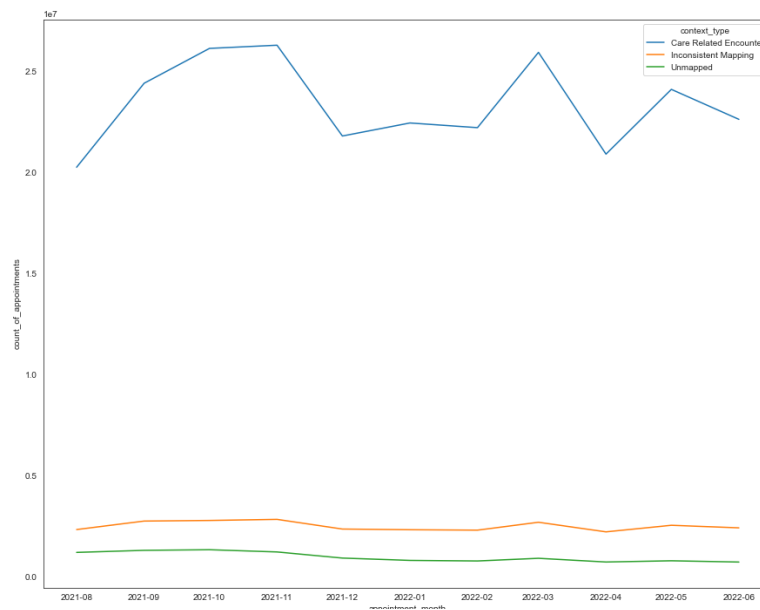


Figure 3: Line plot for count of appointment for context type over time

The context type reveals whether an appointment is a direct patient care encounter, or a non-patient care activity performed as part of patient care. It indicates that patients are directly contacting the NHS through various means.

The various services are further divided into 18 categories considering the scenario. Popular general consultations among them may indicate that patients consult with physicians frequently. With these insights, we can suggest that the NHS focus on these seems to be popular and even help reduce the amount of missed appointments.

Furthermore, we may be interested in looking at the staff level, different healthcare professionals or the time between booking and attending the appointment. This can give more insight into what could be improved. We can determine the number of monthly visits and the average utilisation of services. For this, a new variable utilisation is created based on dividing the sum of monthly appointments by 30. We can plot this as a line graph and get the following:

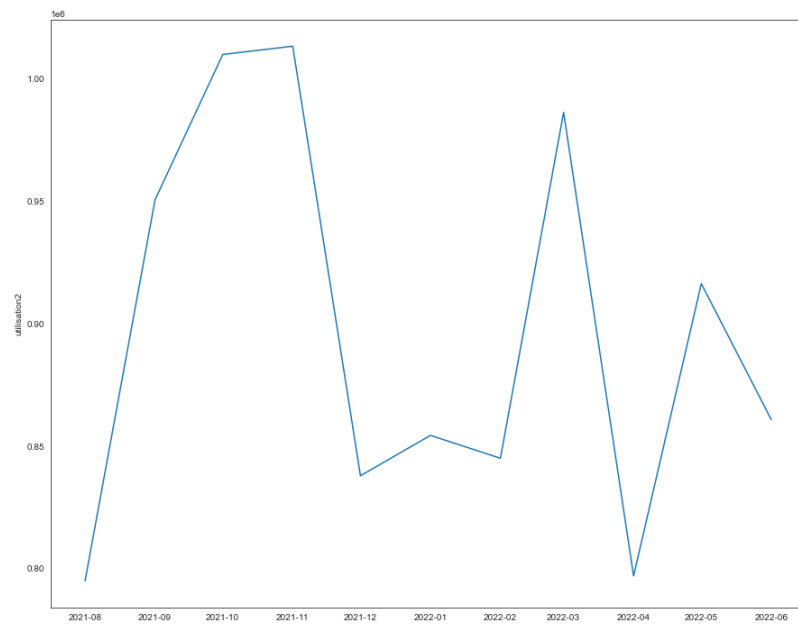


Figure 4: Capacity utilisation over time

We observe that there are fluctuations in which some periods show full or high-capacity utilisation; in other months it is not the case. Therefore, we may need to hire more staff to create a more consistent capacity utilisation. More importantly, depending on different scenarios, we may need to hire more staff.

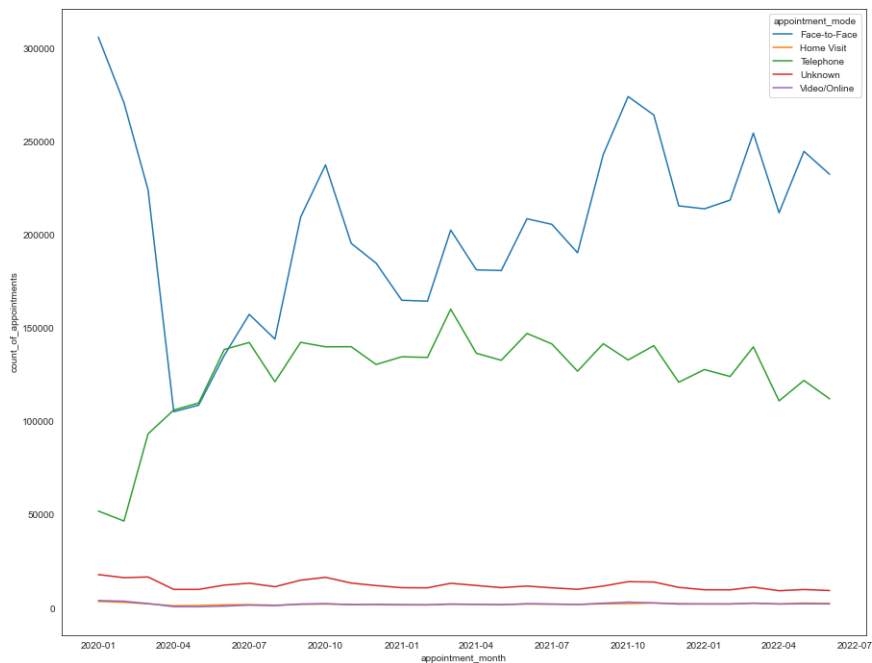


Figure 5: Count of appointments for different appointment modes over time

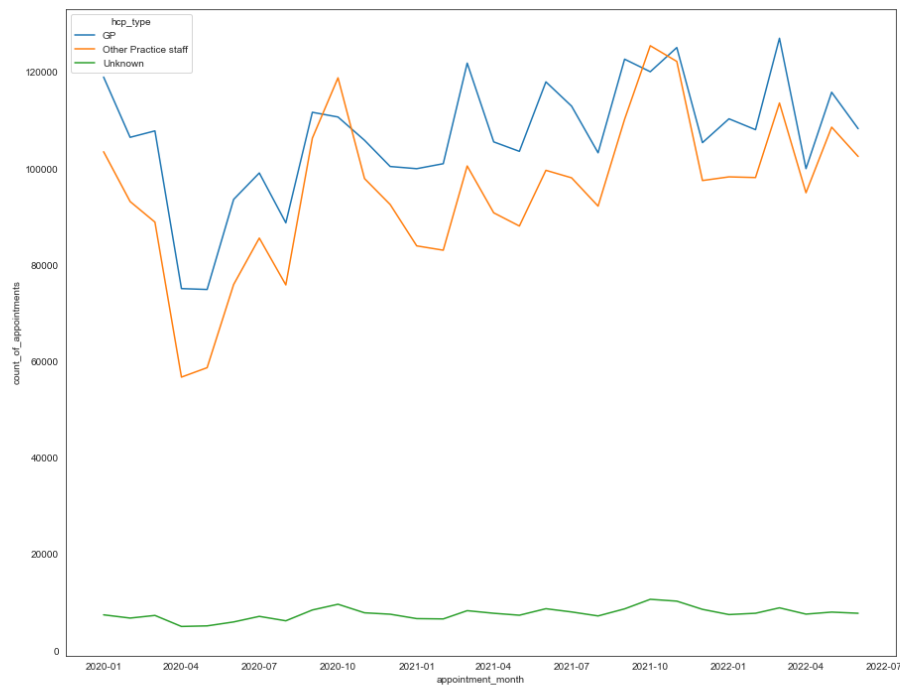


Figure 6: Count of appointments for different healthcare professionals over time

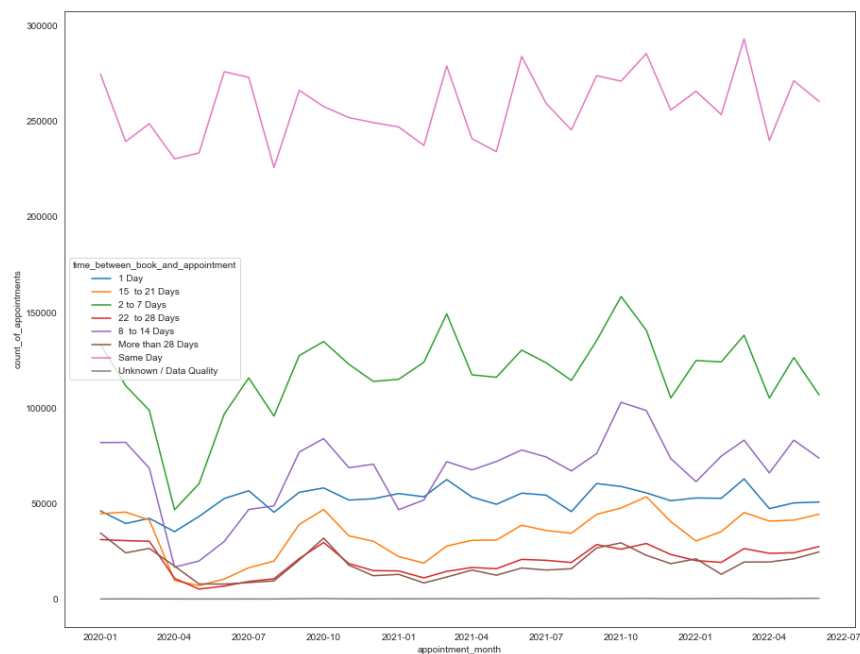


Figure 7: Count of appointments against the time duration of appointments over time

Figures 4, 5 and 6 show different healthcare professions, the mode of appointments and the duration between booking and attending an appointment. These can be significant when tackling the issues of missed appointments. When observing healthcare professionals, we see again that General practices are popular, but other practice staff are popular during some periods. When hiring new workers, they could be nurses, health visitors, counsellors and so on. In addition, by hiring more staff, we can also meet the daily demand for appointments. We would then also follow the most common modes of appointments which are face-to-face or over the telephone.

Additionally, the data gathered via Twitter about the NHS and medicine can be significant. By comparing the lessons learnt and observing:

- Top tweets from a given month in one location
- In your best-performing Tweets, look for trends relating to specific topics.
- When and how often do people interact with tweets
- Monitor key topics or subjects

I examined the most popular tweets, for instance. Figure 7 is of a bar plot that displays the number of the most commonly used tags. The most prevalent one is healthcare. Given this, employing these particular tags may be used in the future to learn more about consumer behaviour and familiar subjects or even to spread the word about NHS or health programmes. For instance, encouraging individuals to get the COVID-19 vaccine by utilising this tag.

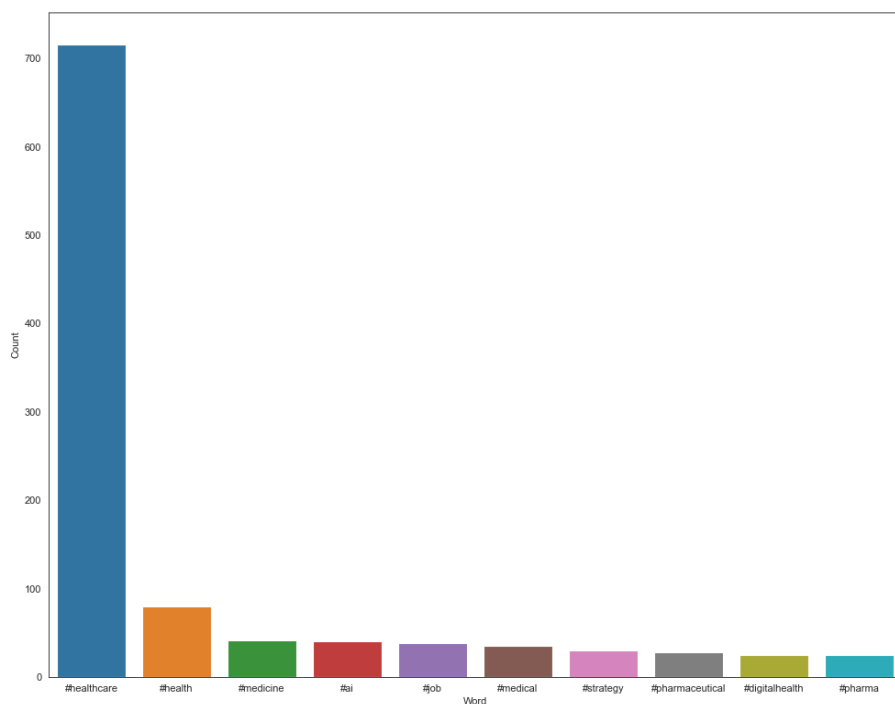


Figure 6: Bar plot for the total count of frequent used tags

GitHub: [https://github.com/Tai3694/-LSE DA NHS analysis](https://github.com/Tai3694/-LSE_DA_NHS_analysis)

Patterns and Predictions

In conclusion, from here, the NHS and the government can learn some insights into why patients are not attending appointments and how to solve this problem. I have created a GitHub repository with all my workings and can be accessed at [https://github.com/Tai3694/-LSE DA NHS analysis](https://github.com/Tai3694/-LSE_DA_NHS_analysis). By examining, manipulating and working with the various available data, we can identify that general practices are the best and most common place for attending appointments.

Following this, we can then consider other essential features such as appointment dates, modes, healthcare professional type, etc. I identified that the capacity utilisation of NHS is now consistent throughout time. They are being utilised at different times, which could be due to insufficient workers. Therefore, I suggest that more staff is needed.

In particular, we can hire staff within general practices but in other forms, such as nurses, counsellors, health visitors, etc. With more teams, we can also meet the daily demand and patient needs as quickly and efficiently as possible. Face-to-face and telephone appointments are the most common modes, so we should stick with these modes and reach patients.

Furthermore, the use of Twitter can aid it in finding specific trends or topics related to healthcare or the NHS. We can see what patients are tweeting, interacting, sharing, and much more. Additionally, we can use this as a platform to communicate and promote specific programmes that will improve the system but also the health of the nation.