NHS Diagnostic Analysis Report

Introduction

In the face of a rise in appointments, the NHS must expand its infrastructure and resources to match. By way of this, the NHS must revise its budget allotment for its locations across the United Kingdom and will need to understand trends in its utilisation of each component in its network to do so. Some stakeholders believe that the current capacity of the NHS is adequate and that efforts to make better use of existing infrastructure and resources are sufficient. This report will investigate various factors surrounding the actual utilisation of resources by the NHS and will provide recommendations based on these.

Analytical approach

For each data source, a data frame was created using the pandas library to derive basic insights such as the presence of null values (using the *.isna()* and *.sum()* functions) and the descriptive statistics of 'count' columns in each data frame (with the *.describe()* function). Descriptive statistics included the standard deviation of the values and the quartiles for the set of values. In the context of the metadata provided, the 'count' columns refer to the number of appointments in a period, whether that date refers to a month or a date each year. The frequency of observations per feature for each data source was also derived, including service settings, locations including top 5 locations, and context types. This was done by taking the count of the number of values per column using the *.value_counts()* and *.count()* functions in tandem for each column within each data source.

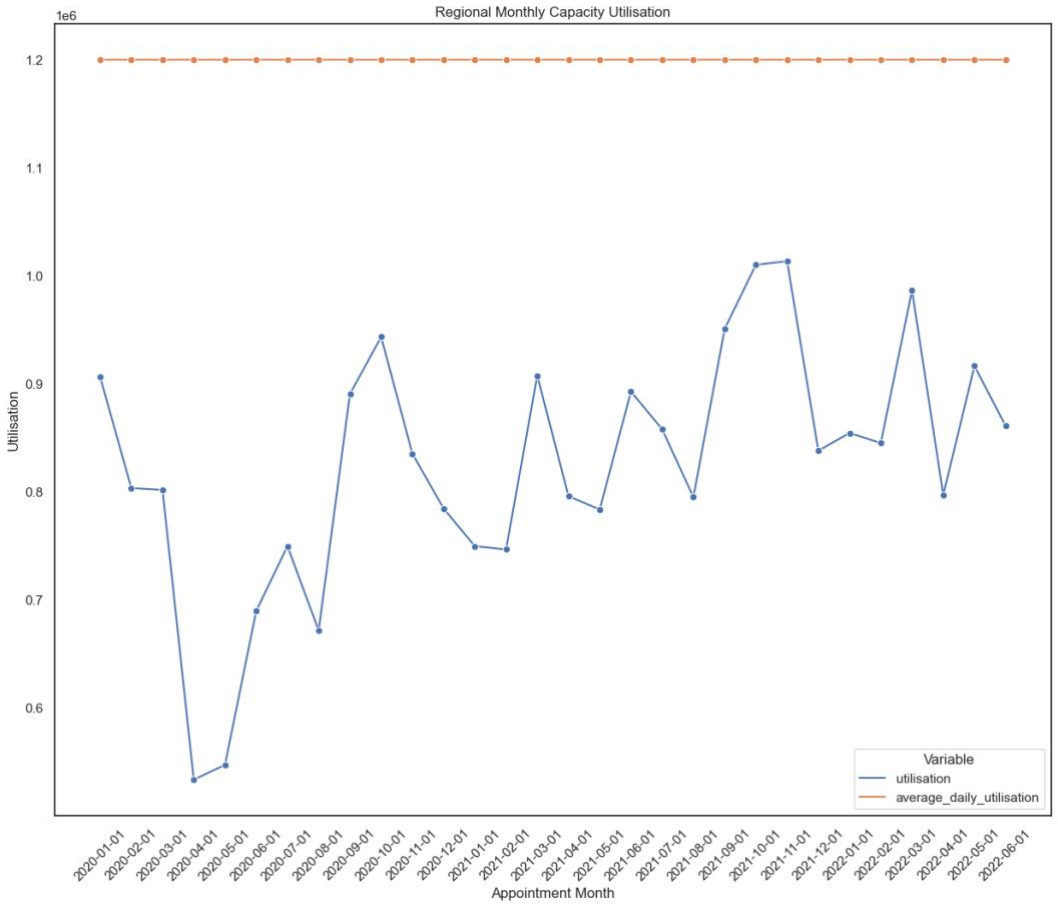
The approach taken to conducting an initial exploration of each data source was required as any null values identified would need to have been replaced or removed as a last resort. This would be done to reduce bias in parameter estimation, manage the representation of samples, and produce valid conclusions derived from each data source.

Observations derived during further analysis include the dates between which appointments were scheduled, and which service settings reported the most appointments in Northwest London for the first 6 months of 2022. For every focal point, the date or month columns were converted from object values to datetime values. From here, a basic function was created to return the minimum and maximum dates for a given data source. Where necessary, subsets or aggregates of data sources were created (with the *.groupby()* function) to filter for relevant features, including appointment date, appointment month, and sub ICB location code.

In terms of the quality and usefulness of the data sources provided, they were robust given that they had been cleaned beforehand for null values and features irrelevant to the client's concerns. The sources could have been more robust in their creation if they spanned across a greater period than 2 years, as among the data sources that contain dates the absolute minimum year is 2020.

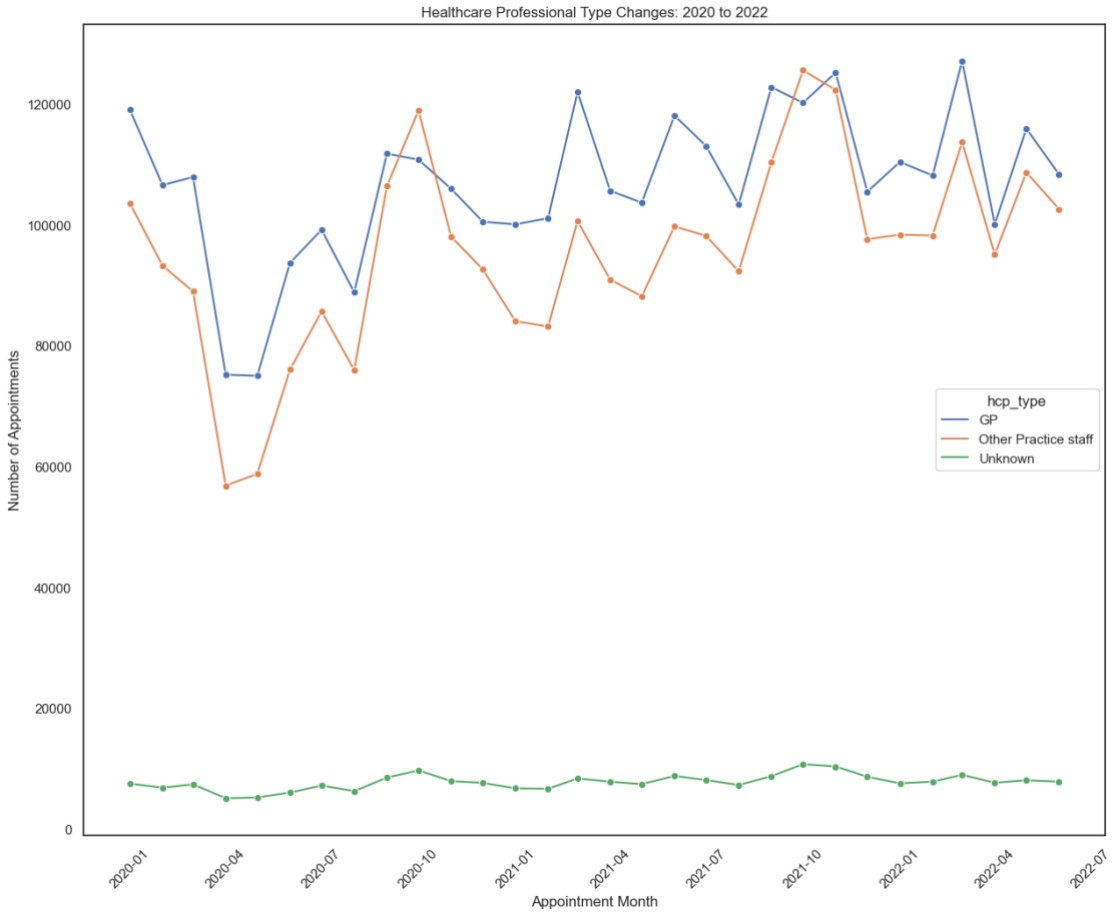
Visualisations, Insights and Recommendations (593 // 550)

Visualisations such as line plots, bar plots and box plots were created to better understand the relationships between relevant features in each data source. Most visualisations are line plots, as these were appropriate for comparing changes over a period, for identifying potential time series trends, and for observing the breakdown of data across sub-groups. With regard to the client's concerns on the actual utilisation of resources, this is generally below average:

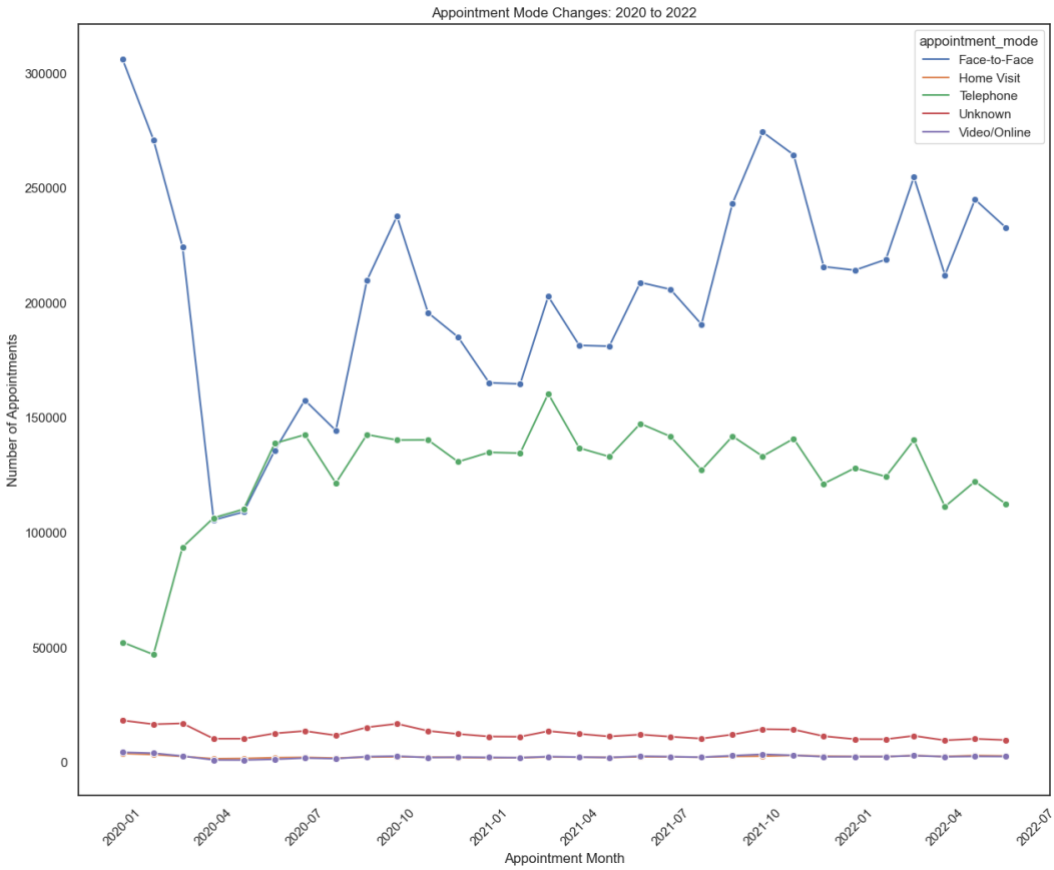


Given that the client provided a figure of an average of 1,200,000 appointments per day, and despite the general increase in appointments per month over two years, the maximum number of appointments for a given month comes to slightly over 1,000,000. The difference of 200,000 appointments between the provisional figure for maximum capacity and the actual maximum capacity of the client between 2020 and 2022 reflects the fact that the NHS should not look to increasing staff levels as a means of managing capacity and should therefore consider making efforts to make better use of existing infrastructure and resources.

Of the resources the NHS possesses and can use more efficiently, focus should be placed on ameliorating the way that General Practitioner appointments are scheduled and conducted given that they are near-consistently the most utilized healthcare professional type:



Further to this, General Practitioner appointments conducted in person, as well as telephone appointments, should be prioritised regarding the improvement of existing infrastructure and resources. This owes to these two appointment modes being the most popular across the time frame that the data sources provided encompass:



Seaborn was used to create a pre-set figure size for all plots, and the hue value for each plot was represented by the number of features in each data source, which was aggregated to focus on relevant features in each source. For example, the aggregated data source for the creation of the line plot for appointment modes was derived using the *.groupby()* function on the Appointments Regional data source with focus on relevant features like appointment month and appointment mode. The aggregated data source was then used as the main data source that Seaborn would draw from to create the relevant line plot, with the hue being appointment mode to visualise any trends or changes in all appointment modes between 2020 and 2022.

Tweets were filtered and aggregated by topics mentioned, first by iterating through each row in the data source containing tweets for hashtags, then by creating a new data set with the *.groupby()* function. The data set reflected the frequency for the discussion of a topic within a tweet, identified by its hashtags. The gathering and analysis of tweets by topic provides insight on general sentiment surrounding topics related to the NHS, such as healthcare. Further analysis could be conducted on sentiment surrounding each topic, which would provide the client with a more consolidated understanding of how its service is received when there are discussions of topics related to them. Understanding that topics surrounding the client's offering are being discussed is useful, though understanding what the public understands the offering to be and the public's experience with the offering would provide more insight into how best to allocate resources.

In summary, it is recommended that the NHS endeavours to make use of existing resources and infrastructure. This is considering the provisional figure for utilisation of capacity and the fact that actual capacity falls below this figure. Given the consistent popularity of the General Practitioner appointments as a healthcare professional type, and the consistent popularity of face to face appointments as well as telephone appointments, the NHS should consider using the General Practitioner resource through both appointment modes in a more efficient manner to begin to maximize the utilisation of their resources.