Background/context of the business scenario

Due to the macroeconomic environment public sectors such as the NHS have been under severe scrutiny in preventing leakages and maintaining cost measures which can be potentially avoidable.

A key component in reducing these costs include the reduction or elimination of missed appointments, this would reap various positive externalities including financial and social.

Due to this investigations into; staffing and capacity in the networks, and utilisation of resources will be scrutinised.

In unpacking the business scenario pythonic code will be used to answer a variety of questions to understand the business problem.

Analytical approach

The data was initially imported into the Jupyter Notebook by way of Pandas, the data was then converted into dataframes which subsequently allowed for easier examination of the data. Calling upon the dataframes to examine and perform rigorous tests allowed understanding of the complexities in the data.

The prerequisite exploratory phase included looking at the descriptive statistics, metadata, structure and integrity of the data (missing and invalid data); this was all completed at a granular level through various pythonic code.

The pythonic code used was varied and it was always ensured that the code was concise and condensed where possible whilst maintaining modularity.

The modular element was important in ensuring each section of the code worked independently, however, could also be structured with other code to perform larger tasks. This also made debugging code easier.

Whilst analysing the data it was clear that there were unlabelled records such as ‘unmapped’ and ‘unknown’, counts were performed on these datasets and it was decided these would be kept as they made up a considerable proportion of the data.

The descriptive statistics were crucial in identifying outliers, this can be observed when the standard deviation was calculated using .describe on the trending twitter hashtags, a formula was then used to call any value which exceeded the standard deviation, those records were then dropped from the dataframe.

As the dataframes shared similar datasets datetime functions were used to ensure that all similar data was synchronised to the same formats for cross examination. For example, appointment month is a shared characteristic, using datetime libraries all dates could be changed to the same datatypes.

Use of other libraries included Seaborn, Matplotlib and Pandas. Personally the use of Plotly would have been a great asset to show visually stunning and descriptive charts however this was not used.

Other functions which were crucial in answering the business objectives included:

- Subsets
- Grouping
- Aggregating
- Sums
- Sorts
- Drops

Visualisations

As part of the business scenario a range of questions were asked to help answer the main project questions. A visualisation was created for each of the questions, this was done through charts, tables and printed written observations by calling defined data.

One of the main visualisations adopted were line graphs. Line graphs are particularly useful in visualising trends in counts or numerical data over a period of time. As much of the analysis was completed using continuous categorical and numerical data such as 'count of appointments' and 'appointment month' line graphs were utilised.

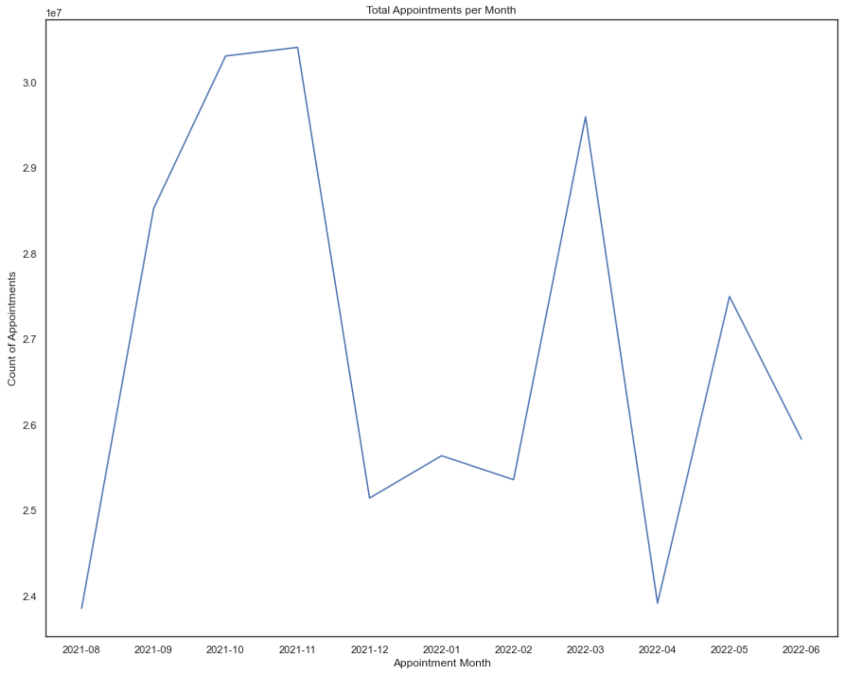
To further improve the visual aesthetic and digestibility of what the data is depicting various functions within seaborn were used to enhance visuals;

- Hue, to drill down into components
- Qualitative and diverging colour palettes helping to represent numerical data with a categorical boundary
- Adjusting position of legend
- Labelling charts and axis
- Changing datatypes to strings/datetime for better readability
- Grouping and aggregating datasets prior to creating charts to ensure buckets of data such as time were easier to read.
- Rotating labels on the axis
- Using escape characters to break-up information

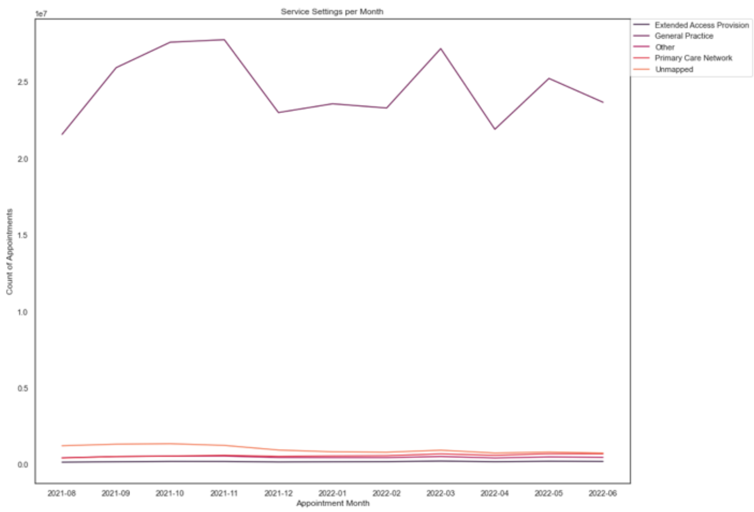
Accessibility of data was also considered, when dataframes were viewed for analysis only necessary data was retrieved. Using syntax such as head, tail, grouping data and dropping unnecessary columns/rows resulted in precise outputs that clearly answered the questions.

Insights

There are in total 106 locations in both AR and NC. As realised in the exploratory phase the dataframes often share results, there was a great scope in the analysis to perform merges and joins to further scrutinise the commonality between datasets to find new insights and patterns.

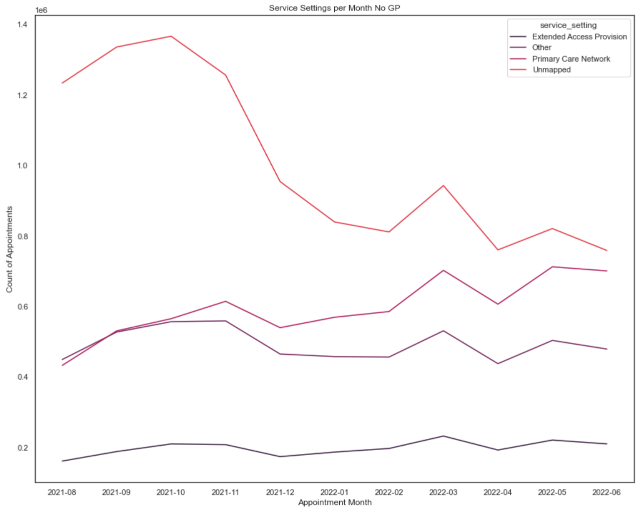


The overall total appointments and appointment month is a good benchmark for understanding the later questions, it shows at an overview the greatest number of appointments in their respective months. We can see around September and October appointments are the greatest and after March and April they are the lowest.



It was clear from prior data exploration that the general practice service setting was the most popular form of appointment, this is useful information in understanding the reasoning for missed appointments.

A separate chart was created to further understand the service settings.

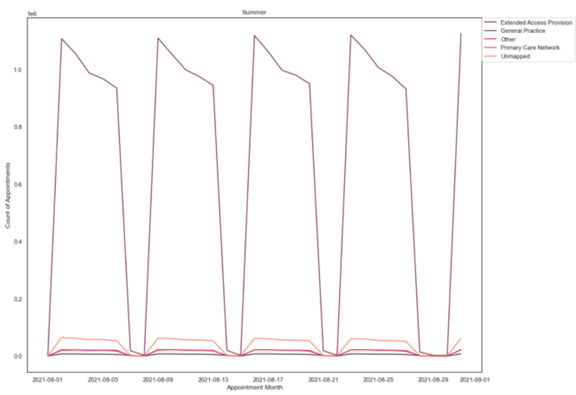


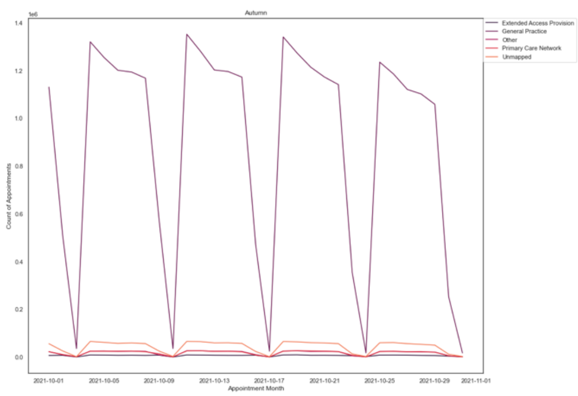
Whilst isolating non GP service settings the trend follows that of the total appointments, the peak in late September and again in March with lower numbers of appointments during December and April.

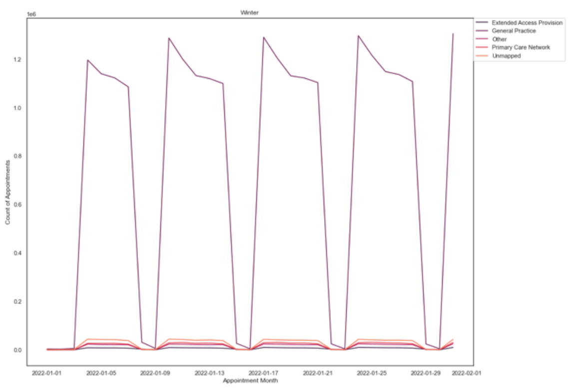
Interestingly the unmapped service setting is the largest service setting after GP, it would be worth investigating this rogue data further as it is so high and at stages doesn't follow expected trends.

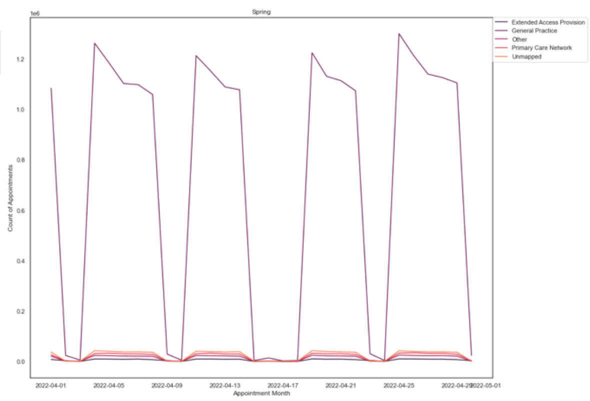
Autumn is the busiest period by the count of appointments, it is worth noting that the majority of appointments are at the start of the week, there is an obvious drop off on the weekends where records reach close to zero. Summer had the lowest count of appointments.

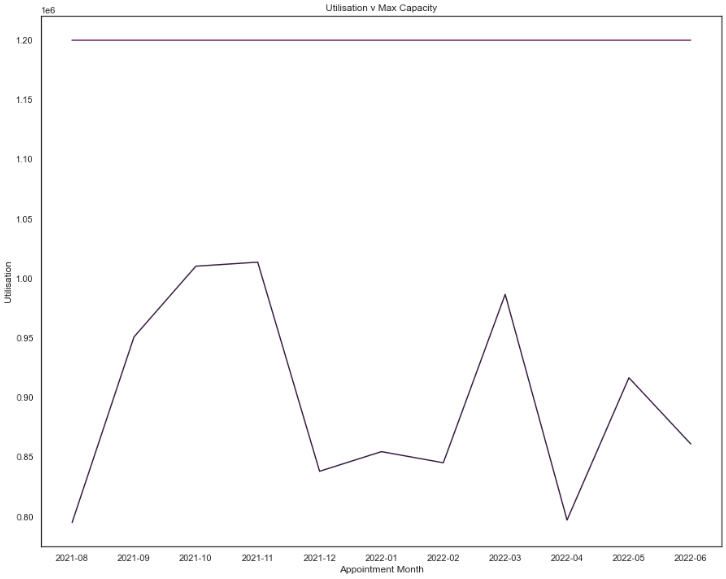
As winter goes on the appointment cases increase this maybe expected with colder temperatures.







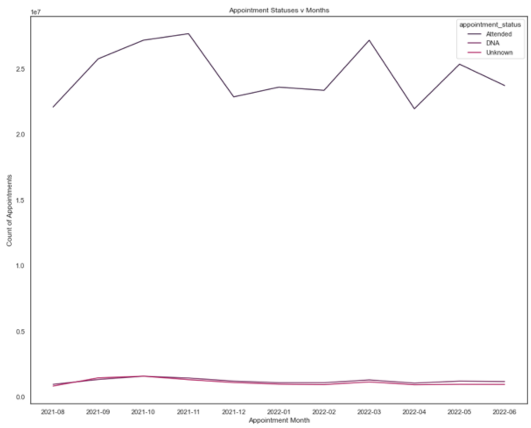


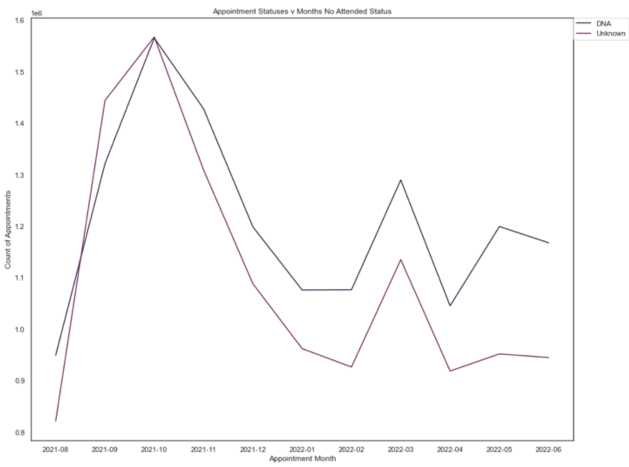


As explained in the first chart much of the charts follow the same trend in time v appointments, we can see here the utilisation follows a similar trend. The maximum capacity lies at 1.2 million, the utilisation never passes the max capacity and even during the busiest months there is still spare capacity. However capacity can include any resource and as we do not have staffing data we cannot confirm that the NHS is adequately staffed.

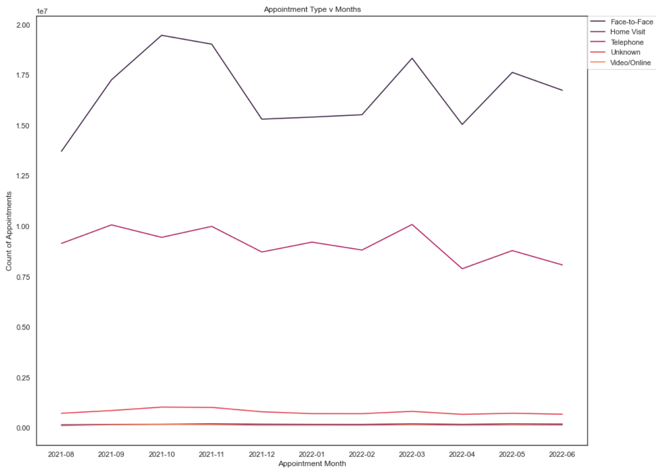
We also do not know how resilient the capacity is to specific areas, such as context types or service settings. Furthermore at stages capacity reaches quite high such as 80%+

Once again appointment statuses follows the first chart trend, more interestingly the busier the period the more likely appointments are attended.

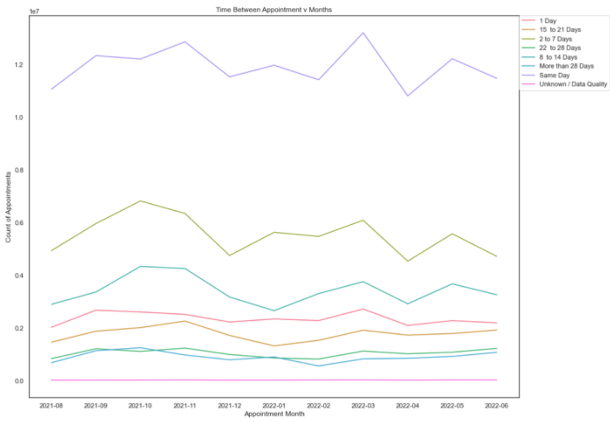




However, this is also true for the inverse and unattended appointments follows the same trend thus somewhat negating one another. The chart on the left excludes overrepresented data in the form of ‘attended’. There is a steep gradient in missed appointments between September and October and again in March. This somewhat confirms the correlation between missed appointments and the busiest periods where capacity is stringent.



Face-to-face interviews are the most popular appointment type followed by telephone. Video and other methods are severely under used despite following the same trends as face-to-face.



The time between booking and appointment is positive, patients can expect to have an appointment on the same day, or within a week.

It would be interesting to see the comparison of staff capacity in relation to the time it takes to book an appointment as although it is easy to get an appointment it may not necessarily mean that the staff capacity or resources are available for it.

Patterns and Predictions

Overall there is a clear trend following the most appointments and the time of year, certain months are more and less busy. This may be due to changing weather conditions and the likelihood of catching illnesses during certain time

It is also generally more busy during the start of the week, this can be attributed to the lack of appointment availability on the weekends, those who require an appointment in the prior week would have to wait until the start of the next week to receive one, therefore resulting in more appointments. If appointments were made available on weekends this could decrease the bottleneck at the start of the week.

There is an obvious demand for General practitioners, and the demand for these members of staff are particularly stringent during Autumn months and at the start of the week. If the staffing could be allocated better to reduce staff numbers during off-peak periods such as summer and during the later stages of the week the resources could be reallocated to ease the pressure on capacity.

The most popular locations are spread across the south of England and the popularity decreases as you travel further north the locations tend to have less appointments. It may be worth considering reallocation of resources in areas with less demand, for example video/telephone consultations can be had with anybody remotely and this would save money and reduce the burden on general practitioners.

Finally data from Twitter has shown that people do not tend to search for NHS and the top tweets are vague in their significance. The interesting hashtags included, jobs and hiring, without context we cannot assume much but may indicate more staff are needed. Finally after some analysis on the tweets it was found that the tweets are global, this means the information is not completely specific to the NHS and the UK.