

Diagnostic Analysis using Python

Context of the business scenario

As part of a team of data analysts, I was required by the NHS (National Health Service), a publicly funded healthcare system in England, to understand the reasons for missed appointments, identify trends and provide insights and suggestions on how to avoid the significant and potentially avoidable costs which it incurs.

As this should be the result of a team collaboration, I've shared the Jupyter Notebook of my analysis on a GitHub repository (https://github.com/michelamallick/LSE_DA_NHS_analysis), where it will be possible for other members to contribute to this analysis and commit changes and updates.

Analytical approach

After importing the data in 3 pandas dataframes and performing a sense-check to verify the presence of missing data with a user defined function, my analytical approach started with an exploratory phase to determine the size, structure and granularity of the 3 tables and assess how to proceed with data cleaning and wrangling.

In terms of quality of the data set, the first evident observation is that a good portion of data in all dataframes is either uncategorised, unmapped or with inconsistent mapping.

Each dataframe covers slightly different periods of time, which overlap between 01/12/2021 and 01/06/2022 at a different level of granularity: while *ad* and *nc* present daily records, *ar* has monthly records. To proceed with the analysis, I harmonised the date/month values to a consistent format (datetime64).

All the dataframes share some common dimensions, but split the data with non-overlapping cross sections, preventing the creation of a meaningful joined dataframe.

Instead, I have subset the relevant dataframes to determine:

- Top five locations, based on number of records
- Most popular service in the top location
- Month with the highest number of appointments
- Total Records per month

And analysed the results with some targeted visualizations:

- Trend of Total appointments by month (in comparison with the top service General Practice)
- Trend of Appointments by Service Setting
- Zoom on Appointment by Service setting: General Practice (top service)
- Trend of Appointments by Context Type
- Trend of Appointments by National Category
- Seasonality of Appointments (Summer '21/ Autumn '21/ Winter '22/ Spring '22)

To complete the analysis of the context, I also analysed the top 30 trending hashtags related to the healthcare in UK.

Finally, to answer to the two main questions:

- Has there been adequate staff and capacity in the networks?
- What was the actual utilisation of resources?

Diagnostic Analysis using Python

I have analysed the following aspects:

- Total appointments aggregated by waiting times for appointments (creating an ordinal variable to visualise the plot in ascending order)
- Waiting times by appointment mode, healthcare professional and attendance status
- Monthly capacity utilisation (considering a capacity of 1.2M appointments)
- Monthly variation of healthcare professional
- Monthly variation of attendance status
- Monthly trends of waiting times
- Spread of service settings

To have a better picture of the context and further the analysis in the right direction, it would have been very beneficial to have data referring to waiting times with location granularity as this would allow to understand where longer waiting times occurs.

Visualisation and insights (350)

After the first exploratory analysis of the data, every step of the analysis has included a visualisation of the data mainly in the form of line graph highlighting the insights of a specific subset. All visualisations have been created with attention of colour blind palettes and adequate scale, in few instances requiring to plot categories separately.

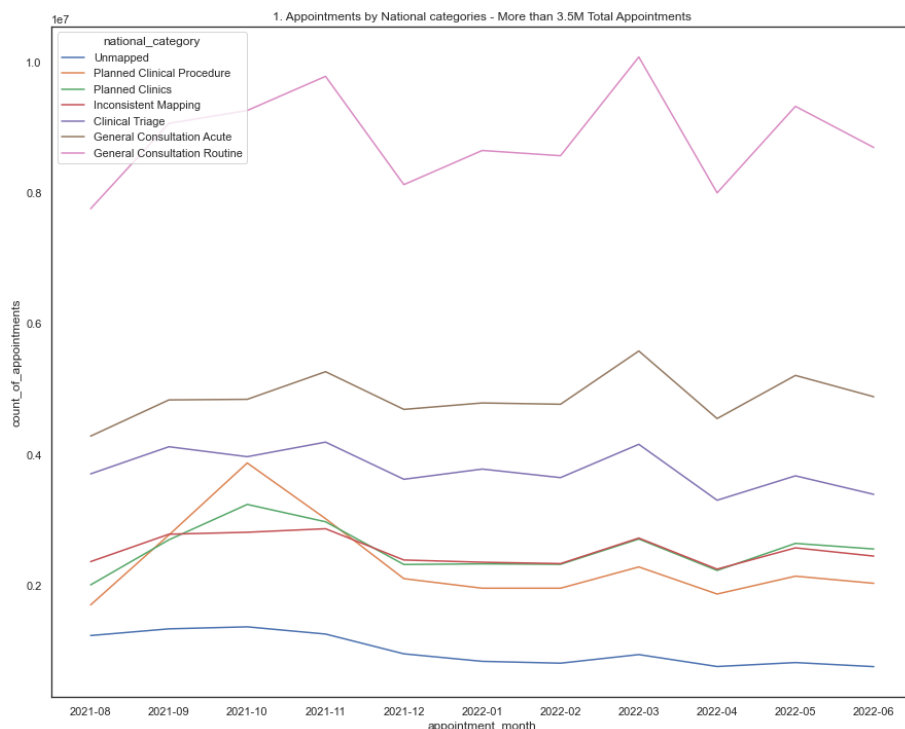
Here's a summary of the insights:

- There are 106 locations, the main service provided is GP (General Practice) as a General Consultation Routine, and most of the appointments are attended
- The three datasets provided cover different periods, overlapping between 01/12/2021 and 01/06/2022. However, the main dataframe used for the analysis (nc - national categories) covers from 01/08/2021 until 30/06/2022



-The **overall trend** of appointments follows the main service, General practice: presents a peak in November 2022, a quiet period during the start 2022, ascending with the start of Spring with a sharp decrease during April (corresponding with Easter) to slightly peak towards the start of summer and a final downward trend

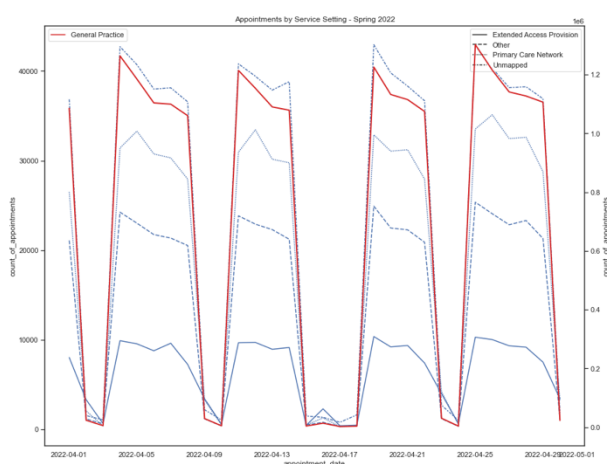
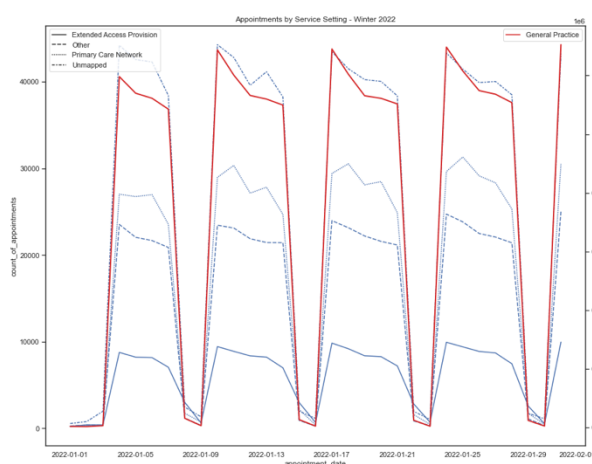
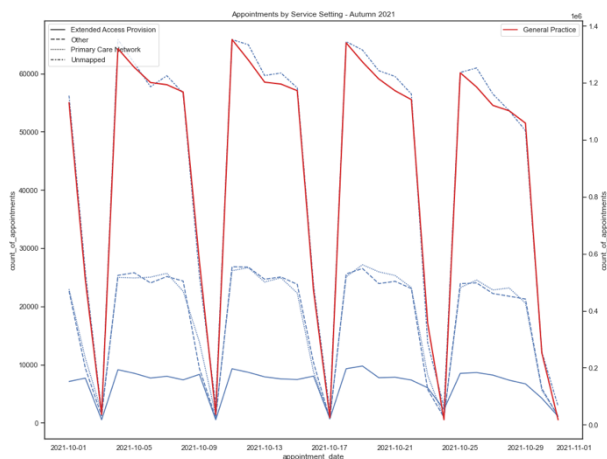
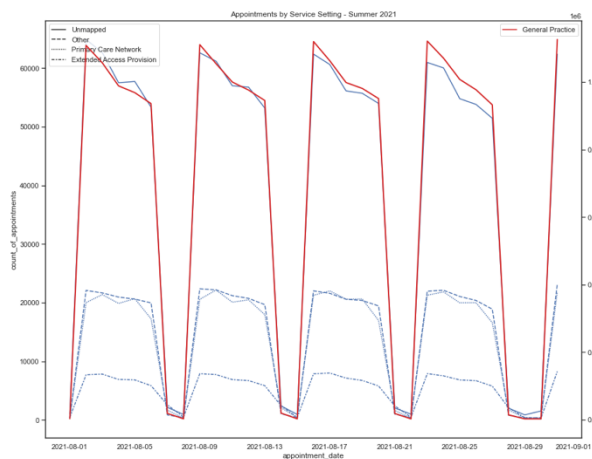
Diagnostic Analysis using Python



- In term of **context**, we only have meaningful data for appointments requested to get care. The rest of the data are either mapped inconsistently or unmapped.

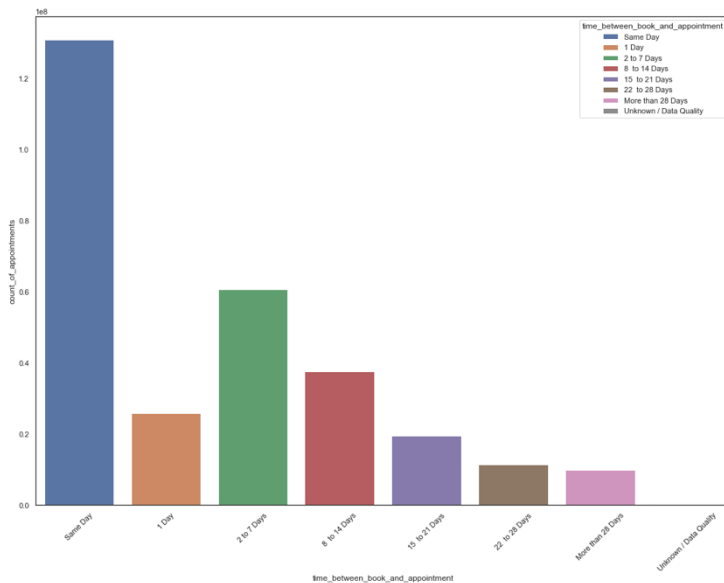
- In terms of **categories**, General Consultation Routine is by far the top category, conserving the trend observed for the previous analysis, with main peaks in Nov 2021 and March 2022. Focusing the analysis on the top 7 categories, is important to notice that a substantial portion of data is Unmapped.

- **Seasonal trends** (Summer '21/ Autumn '21/ Winter '22/ Spring '22) are similar in all seasons, where the services peak at the start at the week (Mondays) with a stable downward trend through the working days and a sharp downward dip on weekends (which extends to bank holidays and festivities), as perhaps the availability of healthcare professionals is limited on Saturdays and absent on Sundays



Diagnostic Analysis using Python

- On the top 30 trending healthcare hashtags of Twitter in UK, NHS does not appear and neither do any of its services.



- Waiting times for appointment seem in total very good, with a majority being same day appointments available. The same can be observed looking at waiting time distributions by appointment modes and healthcare professionals.

[As highlighted with the analysis methods, it would be more meaningful to observe this data with more granularity in terms of locations, as services waiting times might differ considerably.



- Considering a capacity of 1.2M appointments per months, within the timeframe considered the service has been within capacity. The red line highlighted in the graph set the threshold of 80% capacity, which has been breached during peak periods.

Setting aside the quality of the datasets and considerations about accuracy and granularity of the data provided for the analysis:

- The waiting times highlight that though the general capacity of appointments in the overall network is satisfactory; it would be beneficial to investigate the granularity of waiting times based on location, as it is not possible to establish targeted actions on the current dataset
- The utilisation of resources is stretched over 80% in peak periods and is otherwise appropriate. The data granularity for resources and their allocation was very scarce in this dataset. As per the capacity, it is necessary to have a more granular dataset to indicate whose healthcare professionals are available for the different types of services (in this dataset there are only 3 categories)

Diagnostic Analysis using Python

Patterns and predictions

All visualisations highlight a very clear pattern: considering that the appointment requests are driven mainly from GP general consultations, the trend is cyclical and seasonal. Understandably, the health conditions are correlated with the seasons, with a peak at the start of winter (where presumably the cold/flu related conditions are more relevant) and at the start of spring (perhaps for the increase of allergens and pollens and the weather conditions). Requests are significantly low during festivities and breaks (Christmas, Easter, Bank Holidays) and have quite a clear weekly pattern, as described on the insights.

On top of the suggestion provided to improve the quality and the granularity of the dataset for further insights, the resources and capacity planning should focus on seasonality and yearly festivity calendar, focusing especially on the peak periods (yearly, monthly and weekly)

To benefit from the impact of social media for the planning of these services, NHS should involve the public in specific discussions regarding its services expectations, perhaps with tailored hashtags on different topics and services.