

PROJECT REPORT – National Health Service

For the Attention of All Managers and Administrators of NHS Trusts and Hospitals

We were given a task by NHS to find out whether there is enough capacity and subsequent staffing to use that capacity and the resources available and to see if those resources were used to the full extent or there were gaps or shortfalls. This was done in the view of escalating costs of missed appointments at GP clinics and the Government ruling to bring the costs down.

I was given the task of preparing the data files for analysis by making sure they were loaded correctly and that the necessary codes were applied to make it compact enough to be used for creation of data tables and visualisations that will help NHS make decisions or understand where the problems are and direct the thought process on finding solutions. The actual process of preparation is explained separately (**appendix A**).

A quick analysis of the data files shows the sharp contrast between East and South-East of England and the rest of the country. The 4 NHS trusts with the highest number of appointments recorded are in the East and South-East of the country whilst London could only manage one trust. But a substantial chunk of the records of appointments shows the patient booked but did not attend the appointment and a smaller number were recorded as Unknown as it is not clear whether the patient attended or not. These errors can be attributed to diverse ways of recording appointments and statuses, which when collated can give errors like Unknown, making it difficult for analysis and thus arrive at any conclusions. Many records showed up as inconsistent mapping which means that appointments were not recorded in a manner which is standard across all trusts, where some information regarded as important by one trust was not viewed the same way by another trust, thus creating gaps, thus giving way to these errors which cannot be taken into consideration for analysis. (See **Appendix B** for total numbers relating to the errors). These errors in recording show a growing concern for NHS to investigate and rectify the way of recording and managing data and subsequent training of staff to follow standard practices created by NHS.

The visualisation techniques offered by Python and used on the NHS database threw up a few observations which though not many as are still quite important as they shed light on the utilisation of services of NHS and show where the expenditure of funds will be focused and where cutbacks can be made. This has a big ramification for NHS especially in today's times with budgets being slashed everywhere and the government being tight-fisted, NHS will have to manage itself better and perhaps make some difficult decisions in the future.

As can be seen from the visualisations, the number of people attending GP surgeries are the highest and that is the main focus of primary healthcare in UK, and it has been noted that GP surgeries can offer more than just consultation and can act as a community healthcare point where not only consultation but other services like blood tests and treatments for other diseases such as Diabetes can be administered. Healthcare workshops (surgey size allowing) can be conducted, information

regarding hygiene and diseases like Cancer and oral related issues can be communicated to the community. But the visuals also show many cases are not recorded properly which do not give a true picture about patient attendance, which services were used the most and these can have an adverse effect on forecasting of service use, costs involved. This affects the budget, planning of future services and on purchase of drugs needed for treatment. The visuals also shows that most treatments and consultations were done face-to-face and by personal visits from NHS staff. But again, many instances were not recorded properly and have been reflected as unmapped or unknown and these need to be addressed urgently to have a better understanding of which services are needed most and where most money should be spent on. The other visuals highlight what has been seen earlier, sample months taken during each season clearly show GP practices are the focal point of administering primary healthcare.(**Appendix C**).

As can be seen in the plots that GP related services were of the highest demand and face-to-face appointments and care-related appointments (including house visits and visiting old age homes by nurses and qualified carers) are the most in demand and hence it's easy to see where most funds will be spent. But it also shows some trends such as misreporting, incorrect recording of information which causes the results to be skewed and definite conclusions cannot be arrived at. this is expensive to NHS and a drain on the management of NHS as whole. Across all types of services provided by NHS, GP offered services are the most used and hence can be used as a focal point for all NHS communication.(**Appendix D**).

Summary: To summarise all visualisations and data show majority of appointments were made on the same day and were face-to-face, illustrating that patients are going directly to GP or even A&E for treatment and that means pressure on staff to cope with the influx of patients thus increasing waiting times. The NHS was used the most during winter and summer times especially as the pollen season sets in. The visualisations highlight an interesting point where more GP surgeries are needed and they be enhanced to offer services such as blood tests which are still done at many hospitals. But as is seen Hospitals still offer the most type of services and adequate levels of staffing is needed at all times. The dataset also throws light on the fact that a few NHS trusts are recording large numbers of patients whilst others show less numbers which can either be interpreted as some trusts are overwhelmed whilst some are under utilised OR the data is not being recorded properly hence does not give a true picture of utilisation of resources and capacity. No cuts should be done to staffing levels till a standard system of reporting is put in place which will allow the understanding of current capacity and utilisation of resources and services. All visuals point to the fact that patients want a human touch and will attend appointments to consult with the medical staff.

Appendices:

Appendix A

I made sure the data files needed for analysis were imported in to **Jupyter Notebook** through python and its associated libraries such as **Pandas, Numpy** and for visualisations libraries such as **Seaborn, Matplotlib, Requests** and were prepared for analysis by sense-checking using functions like **.dtypes(), .info(), .shape() and .columns**. These functions give information such as names of columns, the data types of each column, the number of rows and columns, and the number of rows with values. Using the code line **DataFrame.head()** gave a view of the data file, thus ensuring it has been imported correctly. This helps prepare the data for calculations and manipulations.

This is the result of the **dataframe.info()** code:-

```
class 'pandas.core.frame.DataFrame'>
RangeIndex: 817394 entries, 0 to 817393
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   appointment_date      817394 non-null object
1   icb_ons_code          817394 non-null object
2   sub_icb_location_name 817394 non-null object
3   service_setting       817394 non-null object
4   context_type          817394 non-null object
5   national_category     817394 non-null object
6   count_of_appointments 817394 non-null int64
7   appointment_month     817394 non-null object
dtypes: int64(1), object(7)
```

I then used functions such as **.value_counts()** which returns the count of occurrences of a particular data item within the column, **.sum()** to find the Totals, to start answering the various questions NHS have asked such as:

Q1 What is the *number of locations, service settings, context types, national categories, and appointment statuses* in the data sets?

Q2 which *service settings* reported the most appointments for a specific period?

I then used functions like **.loc()** to extract the columns needed for analysis, **to_datetime()** function to convert the column into date data type for time based calculations, **.groupby()** function to extract the necessary columns from a larger data set, sort them and extract specific date from those columns.

Examples are:

1) `nc['appointment_month'].value_counts(ascending=True)`

```
2) ac2 = nc_subset.sort_values(['count_of_appointments'],
ascending=False).groupby([nc_subset['appointment_date'].dt.year,
nc_subset['appointment_date'].dt.month]).head()
```

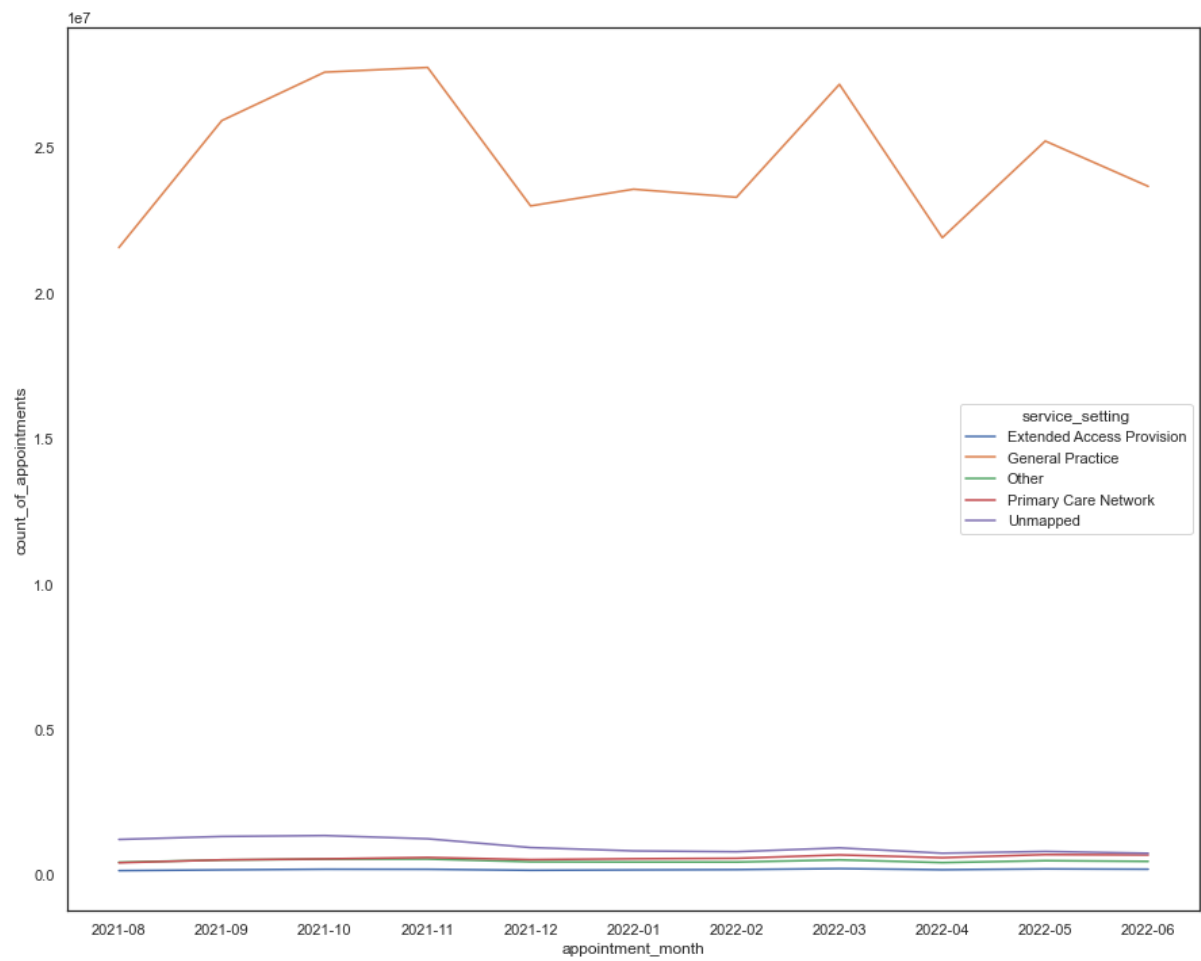
Appendix B

Care Related Encounter 700481
 Inconsistent Mapping 89494
 Unmapped 27419

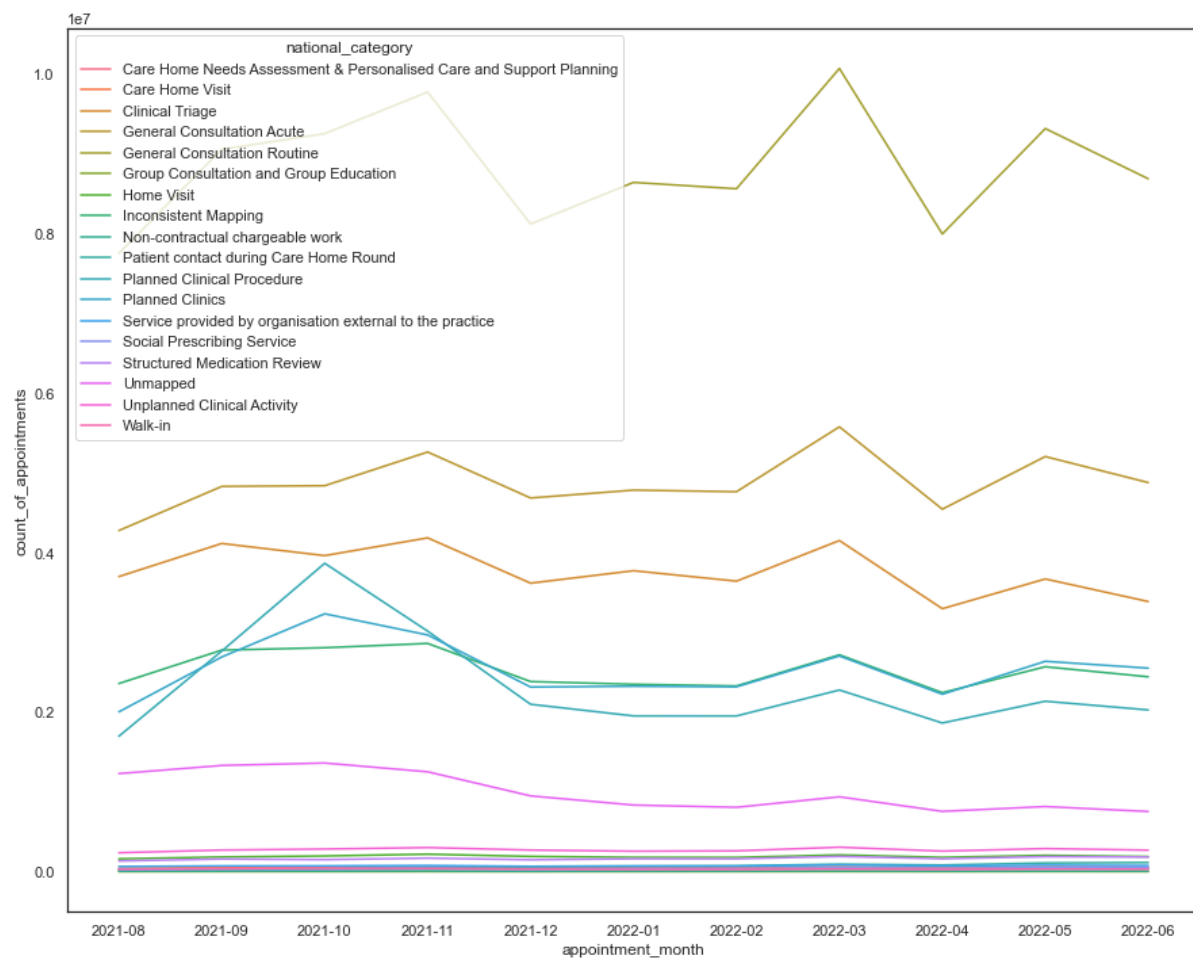
NHS Norfolk and Waveney ICB - 26A',
 NHS Kent and Medway ICB - 91Q',
 NHS North West London ICB - W2U3Z',
 NHS Bedfordshire Luton and Milton Keynes ICB - M1J4Y',
 NHS Greater Manchester ICB - 14L'

nconsistent Mapping	89494	
General Consultation Routine	89329	
General Consultation Acute	84874	
Planned Clinics	76429	
Clinical Triage	74539	
Planned Clinical Procedure	59631	
Structured Medication Review	44467	
Service provided by organisation external to the practice	43095	
Home Visit	41850	
Unplanned Clinical Activity	40415	
Patient contact during Care Home Round	28795	
Unmapped	27419	
Care Home Visit	26644	
Social Prescribing Service	26492	
Care Home Needs Assessment & Personalised Care and Support Planning	23505	
Non-contractual chargeable work	20896	
Walk-in	14179	
Group Consultation and Group Education	5341	

Appendix C

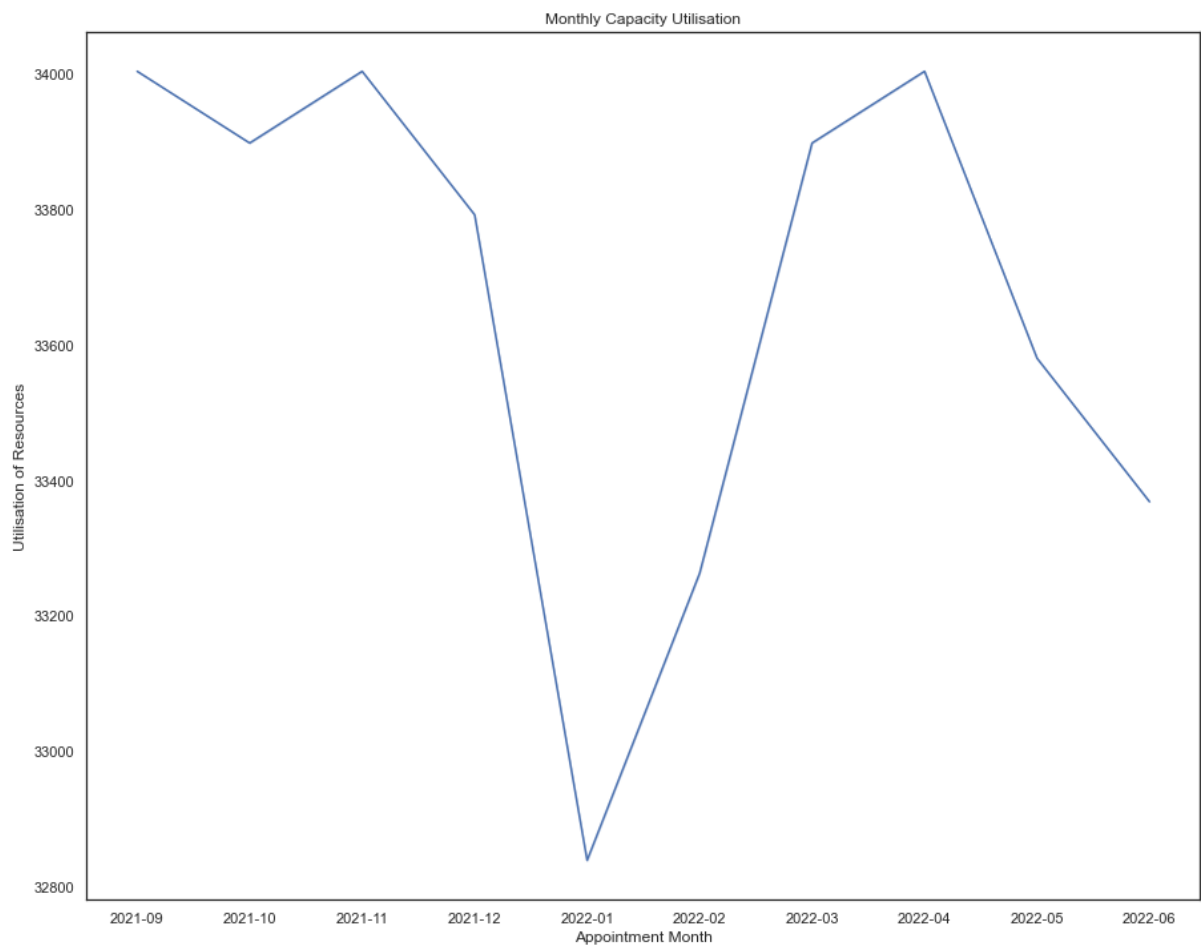


The above visualisation shows the different type of service settings

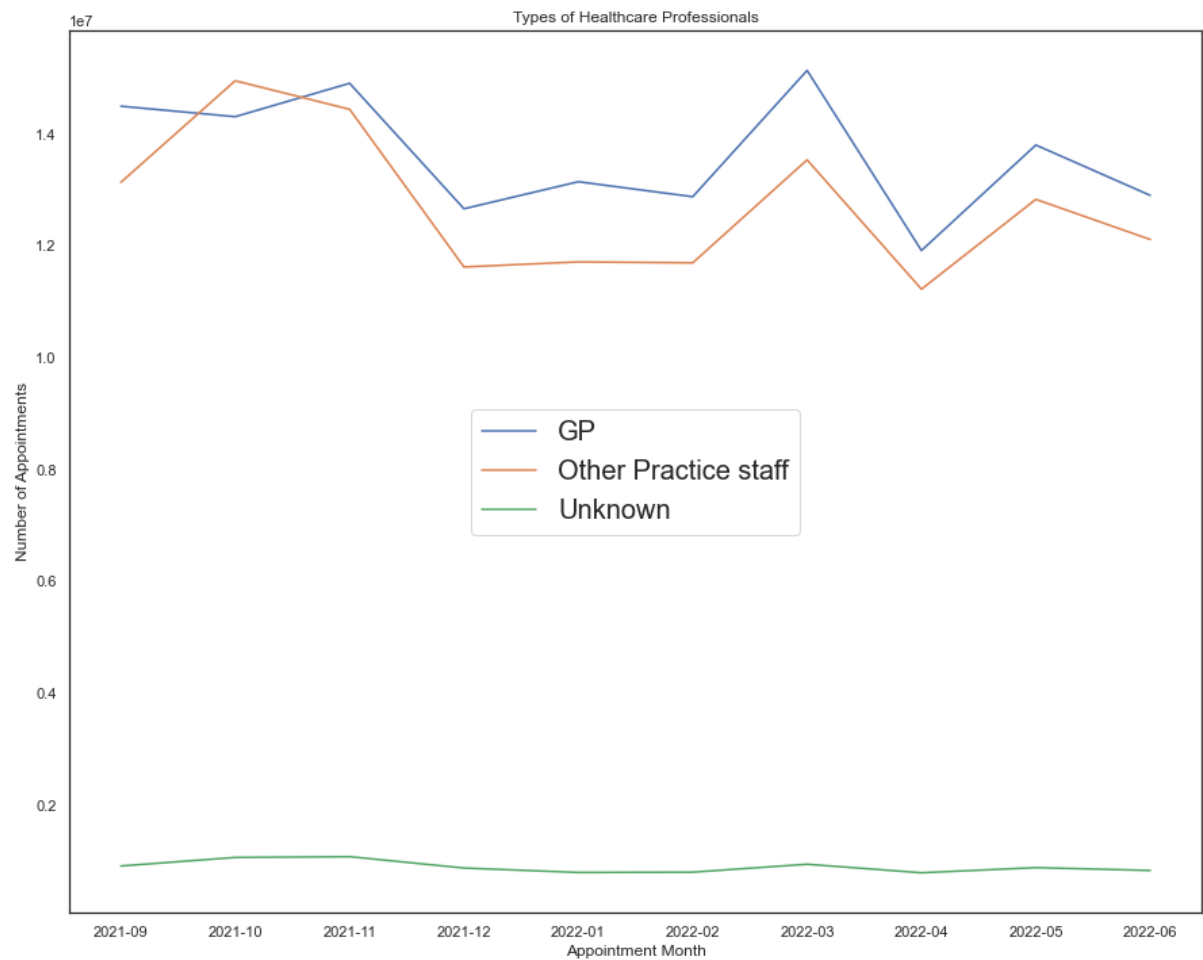


The above visualisation shows different type of services

Appendix D



The above graph shows that NHS was used more in the winter and summer months.



This visual shows where most patients go to for treatment or consultation.



The above visual gives an idea of the type of healthcare professional most consulted by patients