LSE_DA_NHS_Assignment2_Tracy Tsang

This report is to show the analysis details of the initial insights into the NHS appointments issue. The NHS incurs significant, potentially avoidable, costs when patients miss general practitioner (GP) appointments. In order to tackle the problem, it is necessary to identify the trends and factors that may have an impact on the issue.

Using the provided data sets, it is able to analyse the number of appointments breakdown into different categories, such as service settings, context types, and national categories. There are three data sets used as below:

- Actual_Duration – No. of appointments by actual duration and location
- Appointments Regional – No. of appointments by status, mode, hcp type, time between booking and appointment, and location
- National Categories – No. of appointments by service setting, context type and national category.

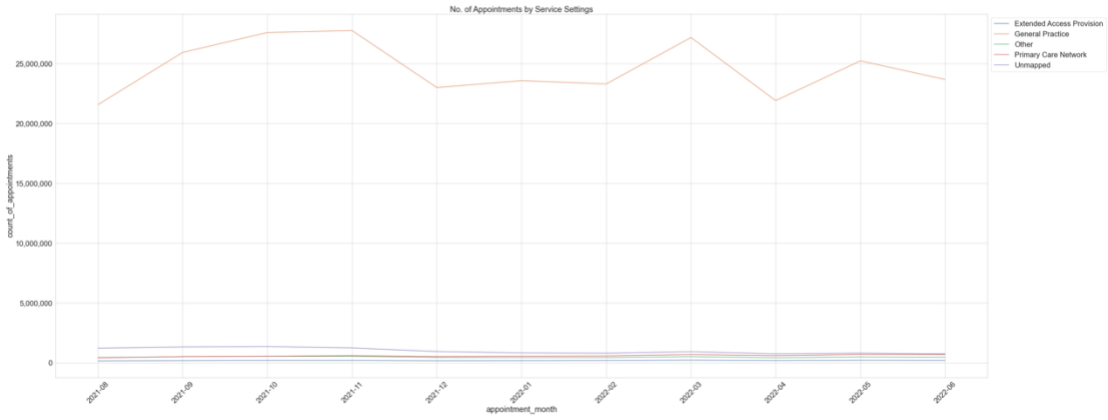
Apart from identifying seasonal trends in no. of appointments made, it will also be worth exploring the top trending hashtags on Twitter related to healthcare and finding the actual utilisation of the NHS resources.

The three data sets are reviewed, and no null values were found. However, the data sets are in different date range and some of the date/time-related data (eg. actual_duration and time_between_book_and_appointment) is stored as text and will require a significant amount of time to convert to numerical values for further analysis. As a start, it's recommended to start with data that is more ready to use, such as "Location_Name", "Service Settings", "Context Type" and "National Categories".

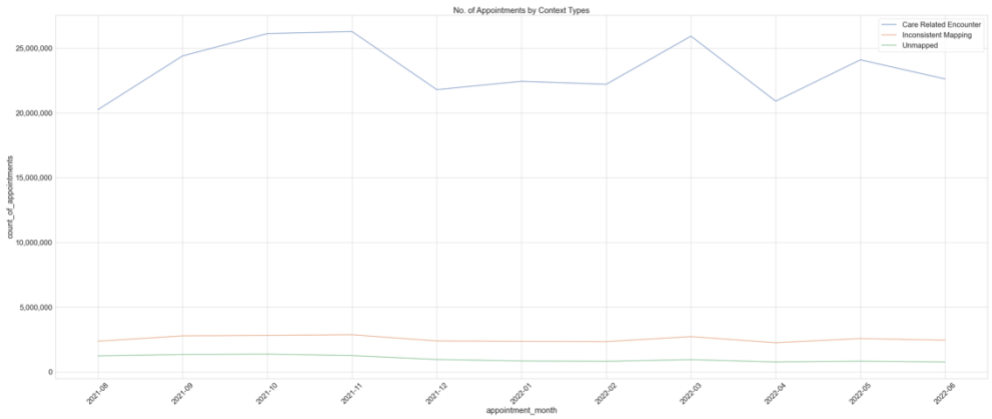
There are 106 locations, 5 service settings, 3 context types and 18 national categories in the data sets. Since the number of locations is particularly large, it may be hard to look for any useful insight at this stage as the result is expected to be highly related to the size of the region and the regional population. To start with, I would prefer to focus on Service Settings, Context Type and National Categories.

Data shows that in general, October and November has the highest number of appointments made. If further breakdown into different categories, below charts shows the monthly trend of appointments made.

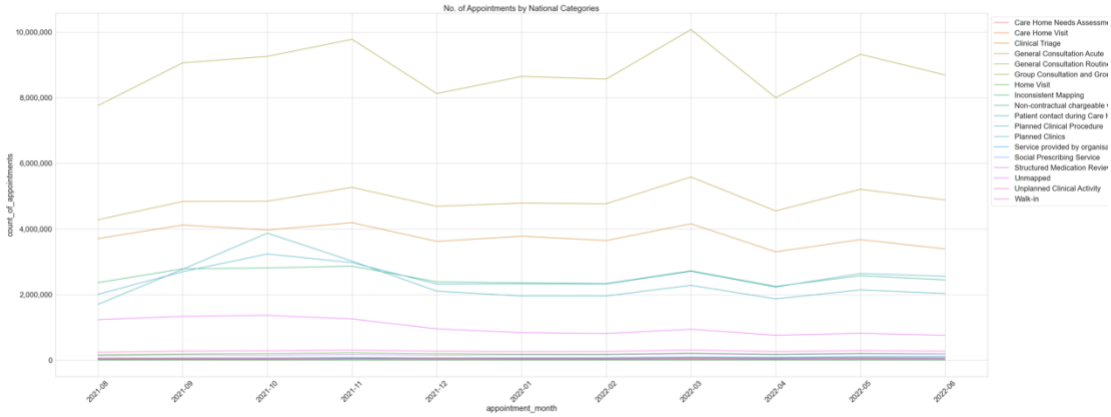
Monthly no. of appointments by Service Settings



Monthly no. of appointments by Context Type

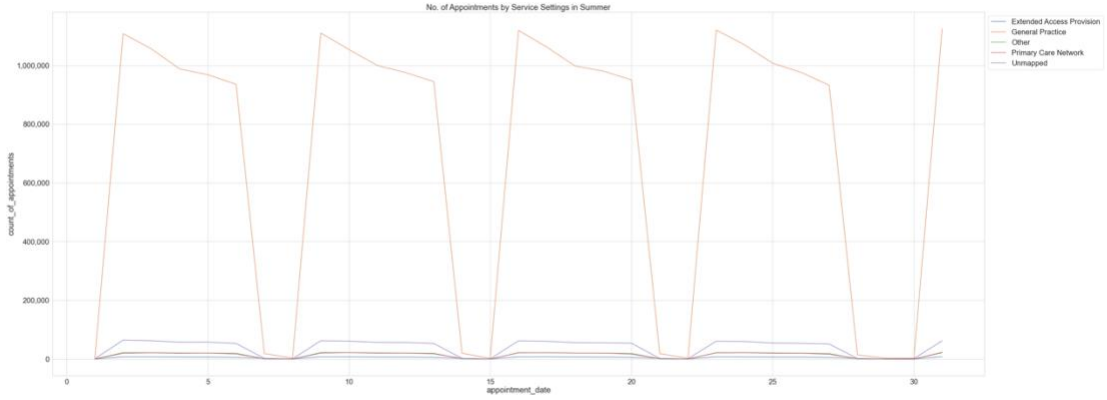


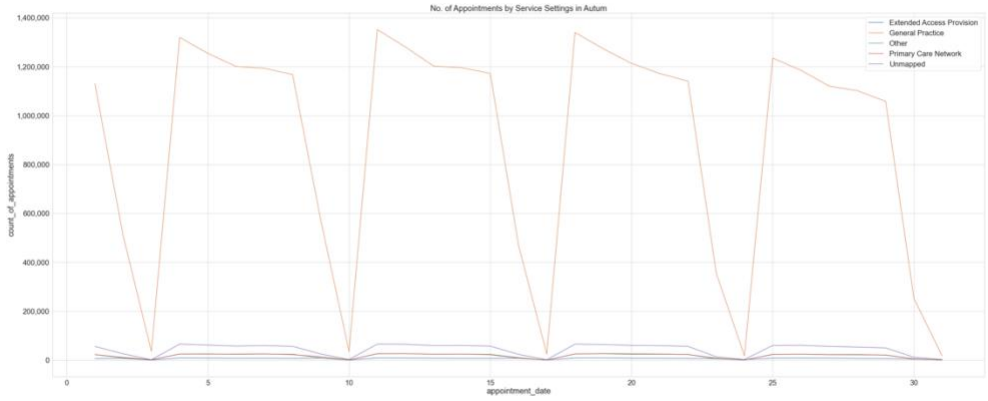
Monthly no. of appointments by National Categories

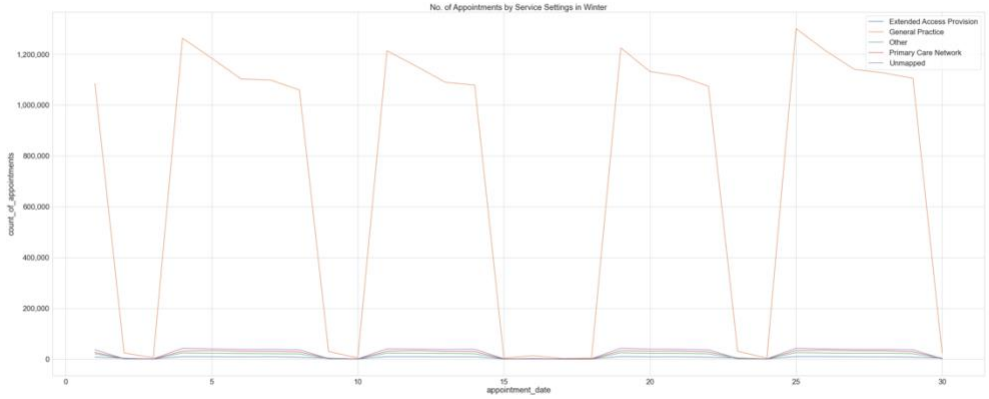


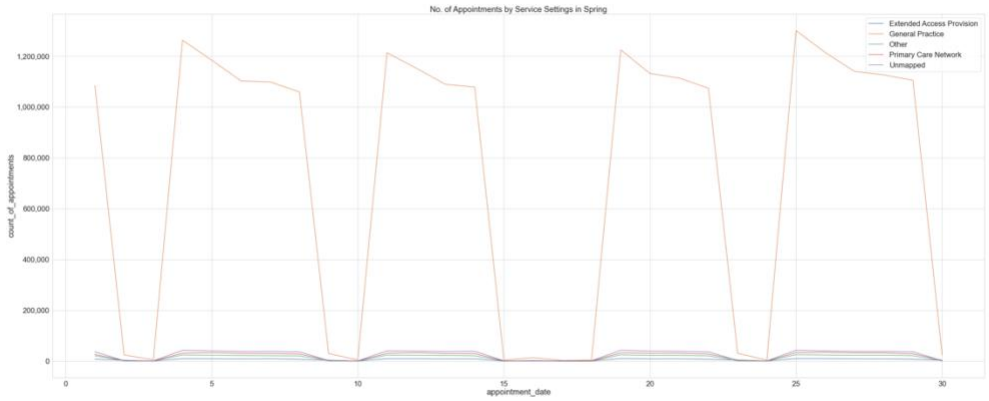
The trends are similar in all three charts, where it reaches its peak in October, November, March, and May, and at its lowest during the holiday season, from December to January.

I then further looked into the daily trends and compare them in different seasons. Four months are assigned to the four seasons as follows:
2021 Aug = Summer
2021 Oct = Autum
2022 Jan = Winter
2022 Apr = Spring


No. of Appointments by Service Settings in Summer
Extended Access Provision
General Practice
Other
Primary Care Network
Unmapped
count_of_appointments
1,000,000
800,000
600,000
400,000
200,000
0
0
5
10
15
20
25
30
appointment_date


No. of Appointments by Service Settings in Autum
Extended Access Provision
General Practice
Other
Primary Care Network
Unmapped
count_of_appointments
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
0
0
5
10
15
20
25
30
appointment_date



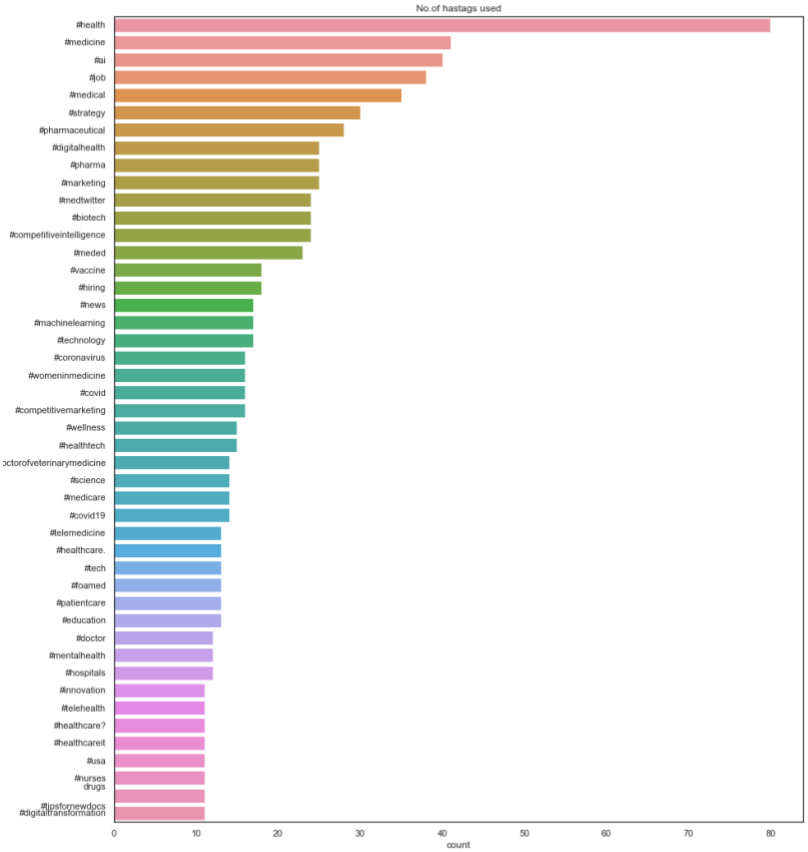


There is obvious weekly trend found in all seasons. This may due to the operation day of services (weekday/weekend) as well as the behaviours of patients.

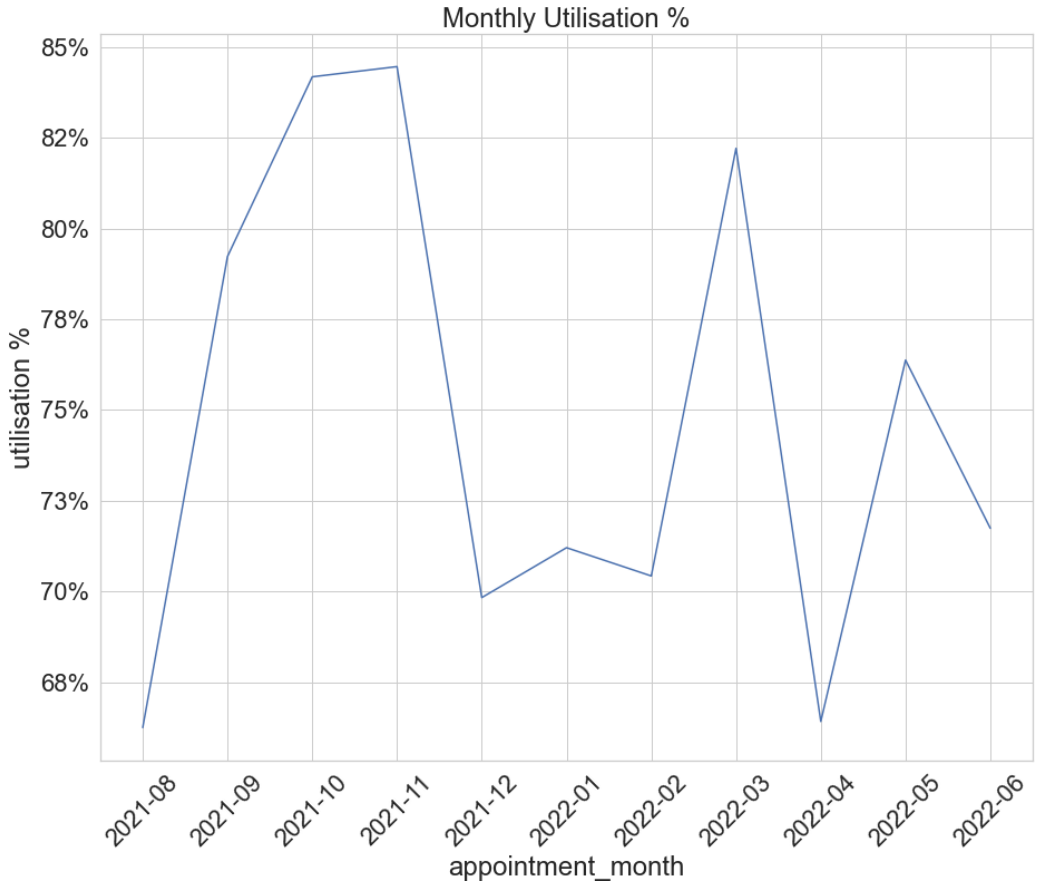
The top trending Twitter hashtag related to healthcare has also been analysed. Out of the 1174 entries in the data set, most of the tweets with healthcare related hashtags are not retweeted nor marked as favourite. 716 entries with hashtag #healthcare, and other hashtags have less that 100 entries found.

	word	count
0	#healthcare	716
1	#health	80
2	#medicine	41
3	#ai	40
4	#job	38
...	...	...
1749	#evestudy	1
1750	#patientdata…	1
1751	#secure	1
1752	#sms	1
1753	\n#csjmu	1

In order to better illustrate the data, the top hashtag #healthcare is omitted in below chart.



At last, monthly utilisation % is found to should the trend of how the NHS resources is used. Calculation is assuming 30days in a month and the maximum number of appointment can be made is 1,200,000 per day.



The graph above is showing that October, November and March are better utilised than in other months with over 80% of utilisation. In December, January and April, probably due to X'mas and Easter holiday, the utilisation is much worse. This could be due to patients cancelling their appointment for the holidays, but further analysis needs to be done to conclude.

This report gives an initial insight into the NHS resources, and shows that there are seasonality factors impacting the number of appointments as well as the utilisation rate. Twitter trending hashtags have also been analysed, but the result is not promising and shows that it may require better marketing or promotion on social media in order to promote NHS services to the population. To further study the problem, I would suggest to investigate the trends in appointment status, the time between booking and appointment and how it relates to the total number of appointments.