NHS CASE STUDY

As part of the National Health Service (NHS) data team my team and I have been tasked to perform a study of the healthcare system. The NHS had been experiencing a great number of no-shows in their medical appointments which suppose for the system significant costs, which could be avoided. The reason behind missed appointments needed to be better understood and investigated.

The government needs a data-informed approach to deciding how best to handle this problem. We conducted a research on the use of resources within the NHS and investigated if the staff and capacity volumes in the networks are adequate, and here we will present my findings.

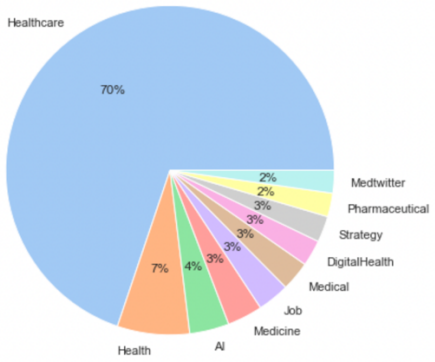
First step in the analysis was understanding the data we were working with and what each database represented. After that first approximation, we proceeded to load and transform the data to be suitable for analysis.

To start, we performed a query to understand the number of locations, the service settings, context types, national categories, and appointment statuses. After that, we identified the date ranges for the different service settings with the following results:

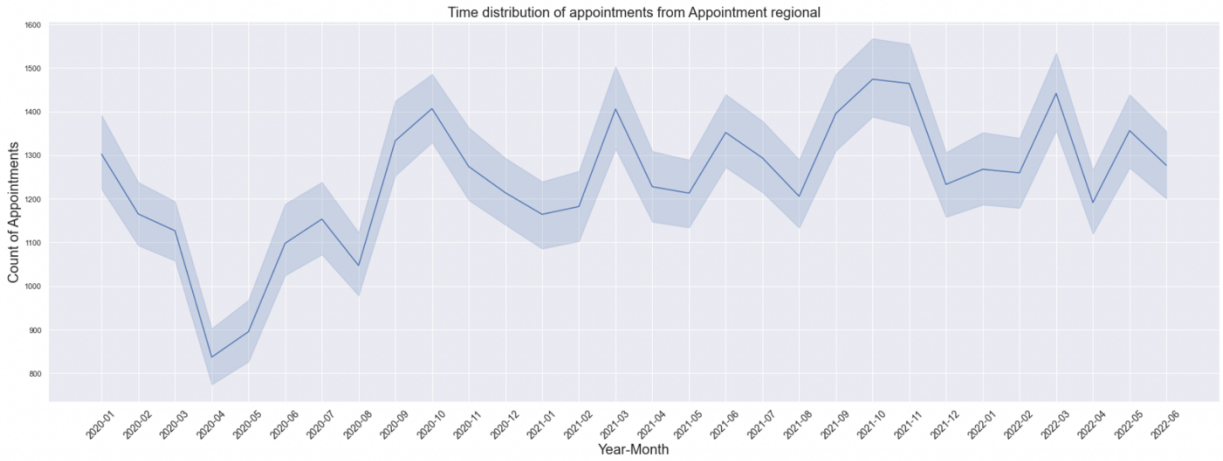
- For National Categories the date range is from 2021-08-01 00:00:00 to 2022-06-30 00:00:00
- For Appointments Regional the date range is from 2020-01 to 2022-06
- For Actual Duration the date range is from 2021-12-01 00:00:00 to 2022-06-30 00:00:00

Following this, we analyzed what the number of appointments and records were per month, with November of 2021 being the one with the largest volume of appointments. The highest peaks among all categories were November 2021 and March 2022.

For the twitter data, we concatenated and transformed the data, as well as eliminate the NA of the database. This yielded the following results: in terms of the worries that the UK population has and expresses on twitter, the top trending hashtags related to healthcare in the UK were #healthcare, followed by #health and #AI. Nothing out of the ordinary.



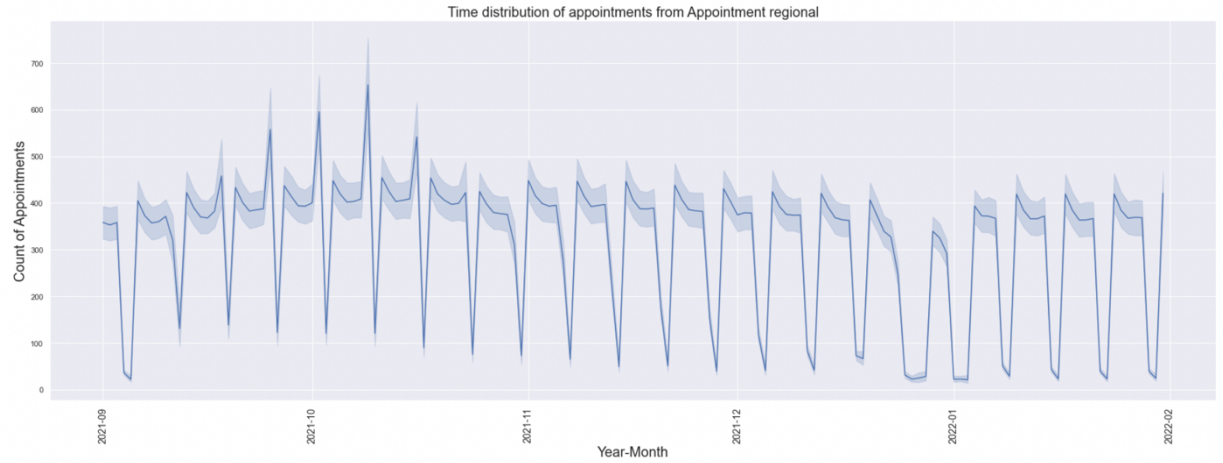
Now, onto the proper analysis. After reviewing the data for some insights, we have found the following: We started using the "Appointment Regional" file, since it is the file with the longest data range, and we can therefore see some temporal patterns more accurately.



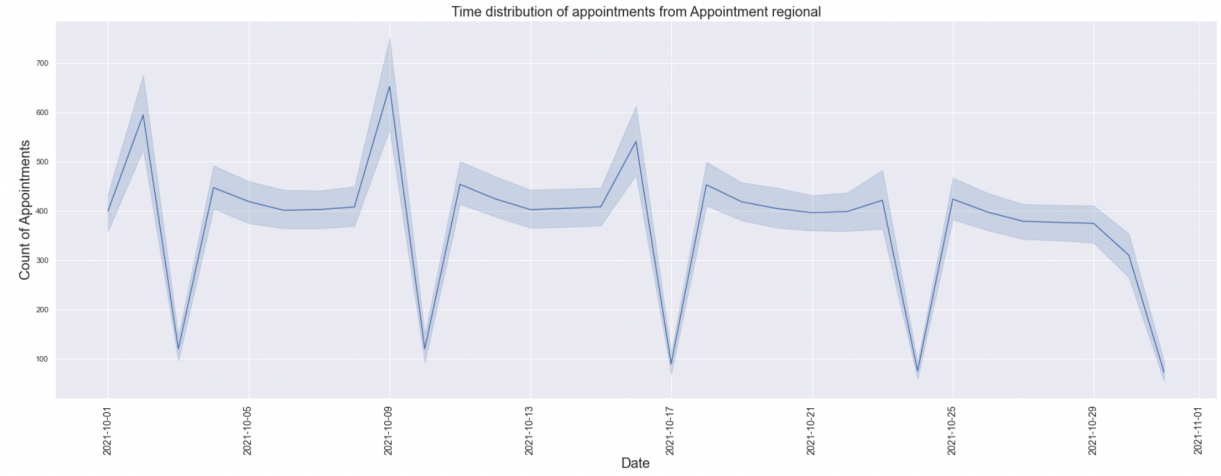
First of all, we can see the effects that the Covid lockdown had between March 2020-June 2020. Even the count of appointments is lower tan usual until October 2020.

If we compare consecutive years month by month, we also can see another patterns. The number of appointments goes up from January to March (maybe flu or other respiratory illnesses?). Then it decreases to a mininum in April, and it rises again in May-June. The maximum is reached in October, and then it decreases again in December (end of the flu, or is it due to Christmas holidays?).

To continue, I take only some months from national categories to analyze day patterns.



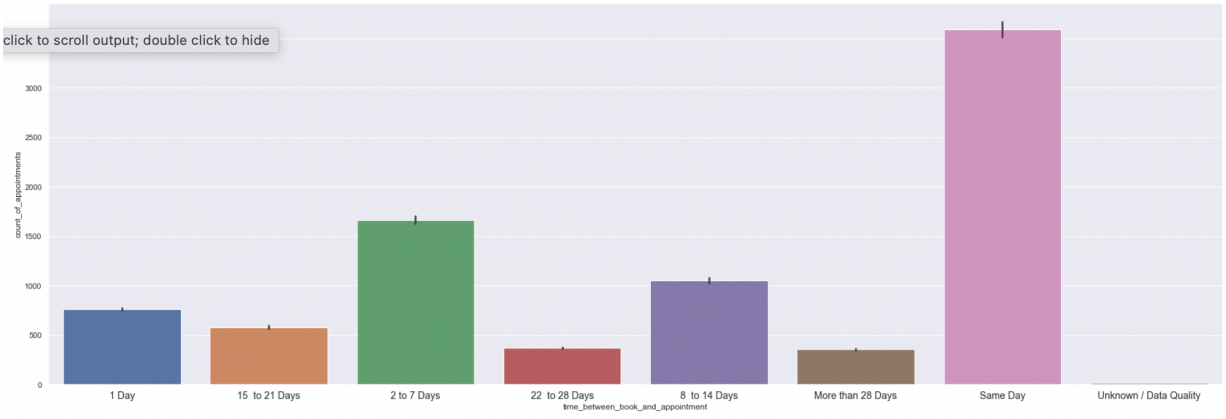
We can see a special repeating pattern. It is repeated peridiocally, except for September 2021 - October 2021.
This pattern is usually due to day of the weeks. We can confirm this if we go deep into some days.



Here we can confirm the weekly pattern. For example on 2021-10-25 to 2021-10-31 We can see a complete week. On Monday, the number of appointments gets to a maximum and then goes down until Sunday.

This is the usual pattern, but we also can see that for some weeks, we can get a maximun at the end of week (on Saturday). For example on Saturday 2021-10-09. This anormal pattern is repeated in September 2021 - October 2021, as it shows in the previous plot with a longer date range.

In terms of the quality of health services, to know some indicators, we can explore and see the gap between the book date and the appointment date. If we see that this gap is big for most of appointments, there could indeed be problem.



After the analysis, it seems there is not a real problem. The most frecuent is Same Day and the second is 2-7 days. But the quality indicators for this measure are not exactly known, therefore no further analysis can be made.

Now we try to find some indicators regarding duration of the appointment. Some quality standards set that the minimum time should be around 15 minutes.

Lets see if we can get some indicators about that. I will make the analysis using ICB location too.
We will drop data with no information of duration.

We will calculate the number of appointments for every duration for every ICB location. We will also calculate the percentage over the total apppintmets in every location, to see if there are some critical locations.

	region_ons_code	sub_icb_location_name	actual_duration	counts_appointment_duration_sum	counts_appointment_duration_total	percentage
0	E40000003	NHS North Central London ICB - 93C	1-5 Minutes	479241	2401937	19.952272
1	E40000003	NHS North Central London ICB - 93C	11-15 Minutes	483885	2401937	20.145616
2	E40000003	NHS North Central London ICB - 93C	16-20 Minutes	331829	2401937	13.815058
3	E40000003	NHS North Central London ICB - 93C	21-30 Minutes	327070	2401937	13.616927
4	E40000003	NHS North Central London ICB - 93C	31-60 Minutes	225123	2401937	9.372561
...	...	...	...	...	...	...
631	E40000012	NHS West Yorkshire ICB - X2C4Y	11-15 Minutes	205961	1025859	20.076931
632	E40000012	NHS West Yorkshire ICB - X2C4Y	16-20 Minutes	123769	1025859	12.064913
633	E40000012	NHS West Yorkshire ICB - X2C4Y	21-30 Minutes	106592	1025859	10.390512
634	E40000012	NHS West Yorkshire ICB - X2C4Y	31-60 Minutes	58408	1025859	5.693570
635	E40000012	NHS West Yorkshire ICB - X2C4Y	6-10 Minutes	281338	1025859	27.424627

We make a plot to see some patterns. First with number of appointments

With this analysis, we are able to see that the most frecuent is 6-10 minutes and 1-5 minutes. The number of appointments is not useful yo know if there is some deviations in a location, because some location can have more patients than others.

So we will use the percentage of appointments of every duration over the total number of appointments for a location.

We can see that 1-5 and 6-10 minutes have high values, so in some location the appointments around 1-10 minutes can be more than 50%, that is not a good indicator.

We will try to obtain the locations with more percentage for 1-10 minutes duration. We will also calculate the mean of percentage for all locations.

The mean of percentage of duration 1-10 is: 48.69604219906763

The number of locations with more percentage of 1-10 minutes appointments is: 58, these will need to be further investigated.