Executive Summary

The National Health Services (NHS), a publicly funded healthcare system in England, incurs significant, potentially avoidable, costs when patients miss general practitioner (GP) appointments. The reasons for missed appointments need to be better understood as explained by The British Medical Association (BMA) chair Professor Philip Banfield:

While it is frustrating when patients do not attend, the reasons why this happens should be investigated rather than simply resorting to punishing them. Financially penalising patients inevitably impact the poorest and most vulnerable in the community (GP Practice News 2022).

NHS's Major Business Questions:

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

Approaches to support the answer analyzing data given from different aspects:

A. Data Source Exploration	*What is the number of locations, service settings, context types, national categories, and appointment statuses in the data sets?* *What is the date range of the provided data sets, and which service settings reported the most appointments for a specific period?* *What is the number of appointments and records per month?*
B. Relational analysis between individual definitions of data	*Were there adequate staff and capacity in the networks?* *What was the actual utilization of resources?*
C. Trends and Patterns	*What monthly and seasonal trends are evident, based on the number of appointments for service settings, context types, and national categories?* *What are the top trending hashtags (#) on Twitter related to healthcare in the UK?*
D. Insights and Explanations	*What possible recommendations does the data provide for the NHS?*

A. Data Source Exploration

There are 4 files provided by NHS and we import all of them and check the basic information together. Three of the files were obtained from NHS internal system but the source of the tweets file should be considered an external data source which should be used with care.

Data File Time Frame

National Category- nc	2021-08-01 00:00:00 to 2022-06-30 00:00:00.
Actual Duration – ad	2021-12-01 00:00:00 to 2022-06-30 00:00:00
Appointment Regional	month started from 2020-01 to 2022-06

Data frame Summary:

1. The 3 dataframes' earliest entries were all different to each other.
2. The earliest record of information appeared in the regional appointment dataset.
3. The 3 dataframes had the same record of timeline for the appointment records.
4. There were no obvious master key id common and unique within all files, so that the files were highly possible to be extracted from independent systems.

 # How many locations are there in the data set?
There are 42 locations in the data set.

 # What are the five locations with the highest number of records?
The 5 location with highest number of records

icb_ons_code	count_of_appointments
E54000050	16882235
E54000054	14358371
E54000057	13857900
E54000008	13250311
E54000027	12142390

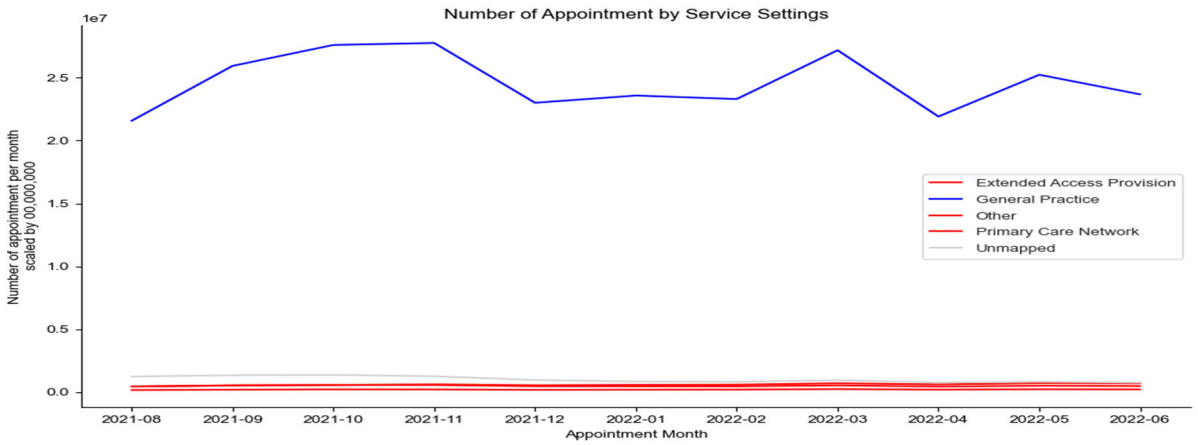
Data Type Check:

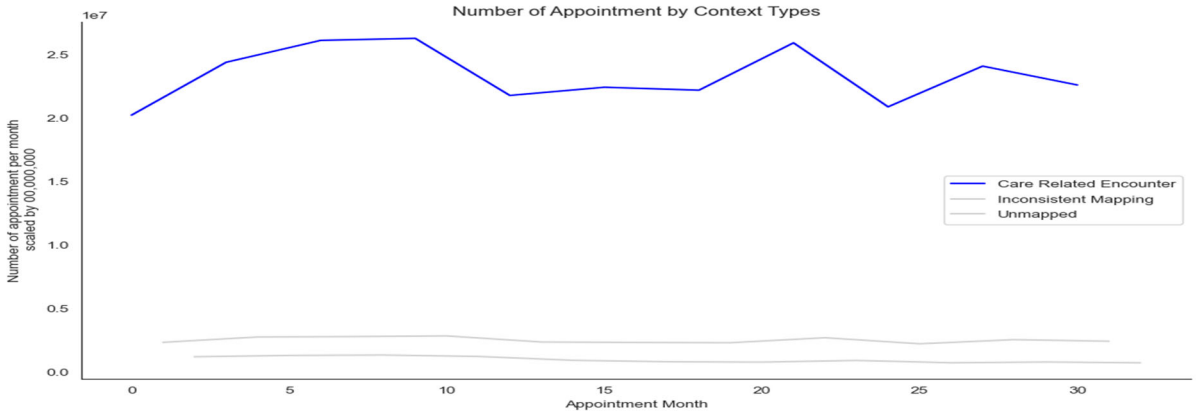
In the ad file, the actual duration was categorized into several groups. In fact, it could be of help to estimate the time spent with each appointment visit. We could use the group tag, e.g. 31-60 minutes, to give an estimate of 45 minutes actual parameter for that particular appointment.

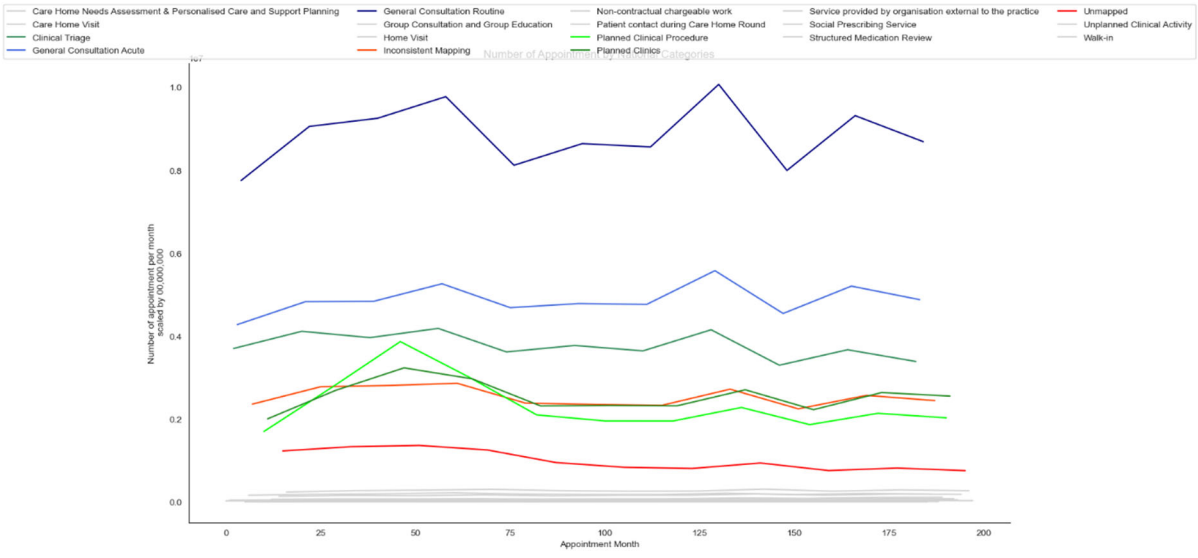
As such, by multiplying the daily count of appointments, we are able to give an average manhour on that day

Trends and Patterns

1. Create three visualisations indicating the number of appointments per month for service settings, context types, and national categories.¶







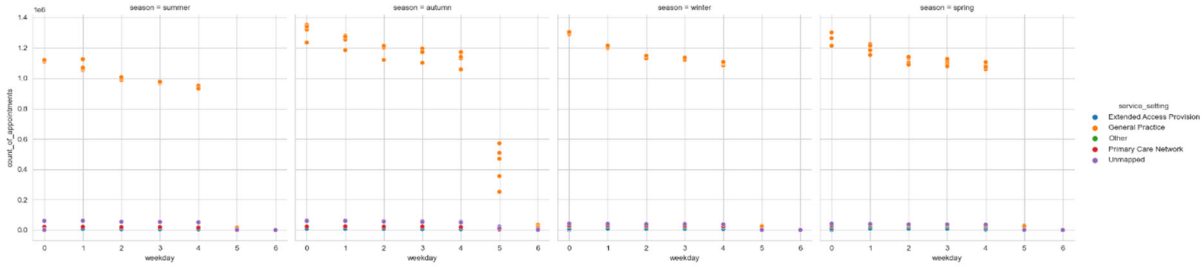
INSIGHTS FROM THE VISUALIZATION

1. The overall trend for the service_setting and context types were moving with very close and similar pattern which were likely that the input meta data were realized as the same from the practitioners.

2. For the national category, we could identify three major zone by observation:
    - the Red Zone (Unmapped, Inconsistent Mapping)
      the "inconsistent mapping" was moving with different pattern
    - the Green Zone(clinical triage, planned clinical procedure & planned clinics)
      "Planned Clinical Procedure" appeared some abnormal pattern in the period from 2021-8 to 2021-12 thus advised to investigate the roots.
    - the Blue Zone (General Consultation Acute, General Consultation Routine)
      This appeared to have same moving pattern as with the service setting and context types.

3. FORECASTING MODEL COULD BE SUGGESTED based on NATIONAL_CATEGORIES AS CONTRIBUTING FACTORS for resource allocations.

2. Create four visualizations indicating the number of appointments for service setting per season. The seasons are summer (August 2021), autumn (October 2021), winter (January 2022), and spring (April 2022).

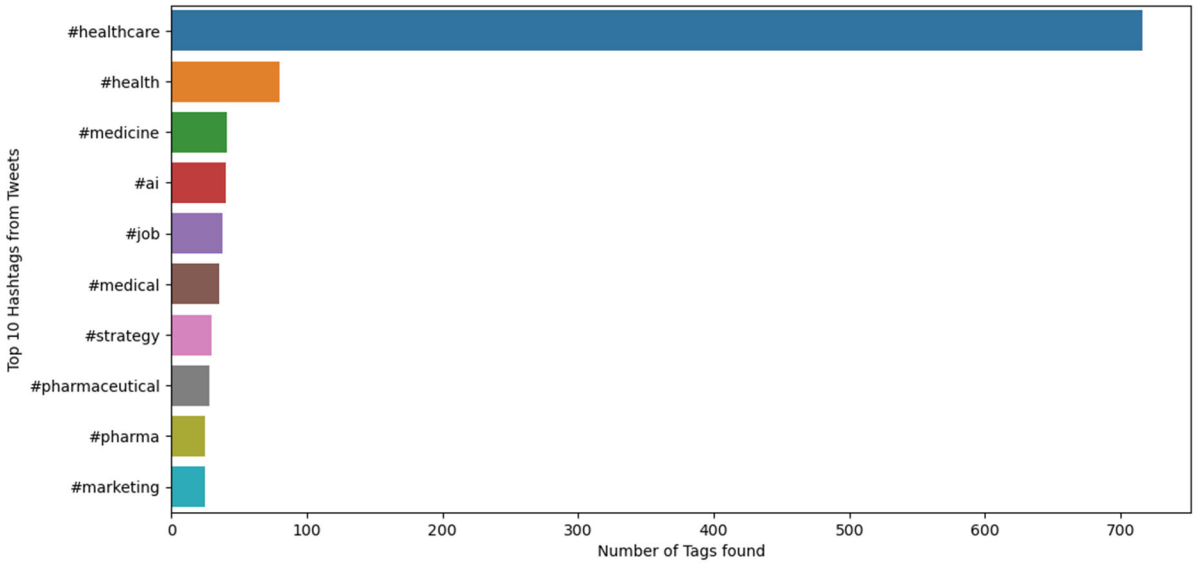


INSIGHT FROM THE SEASONAL PLOT¶

1. Patients were generally not likely to have appointment on Sat and Sun, but most likely on Mon, except the Sat in AUTUMN
2. Appointments would be less during the SUMMER time in general on average

ANALYZING EXTERNAL DATA SOURCE

Tweeter is a well-known social network platform and millions of user expressing their thought and idea on every dimensions of life. We are given the scraped data from the tweets data files and we are going to explore the message from the public community with a superficial approach. It is always to consider the related legislation for the purpose of using the data and how the privacy of users could be maximally secured.

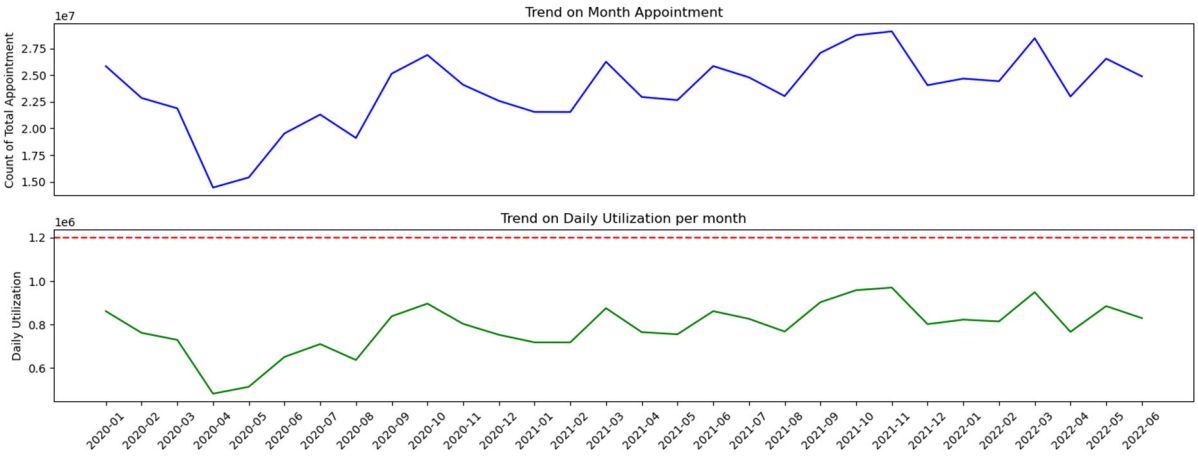


INSIGHTS FROM THE TWEETS

According to the hashtag search from users' message, the healthcare tag was found exceptionally high compared to other tags in the tweets. That could mean substantial concerns about the healthcare related issues during the studied period.

BUSINESS QUESTION:

Should the NHS start looking at increasing staff levels?



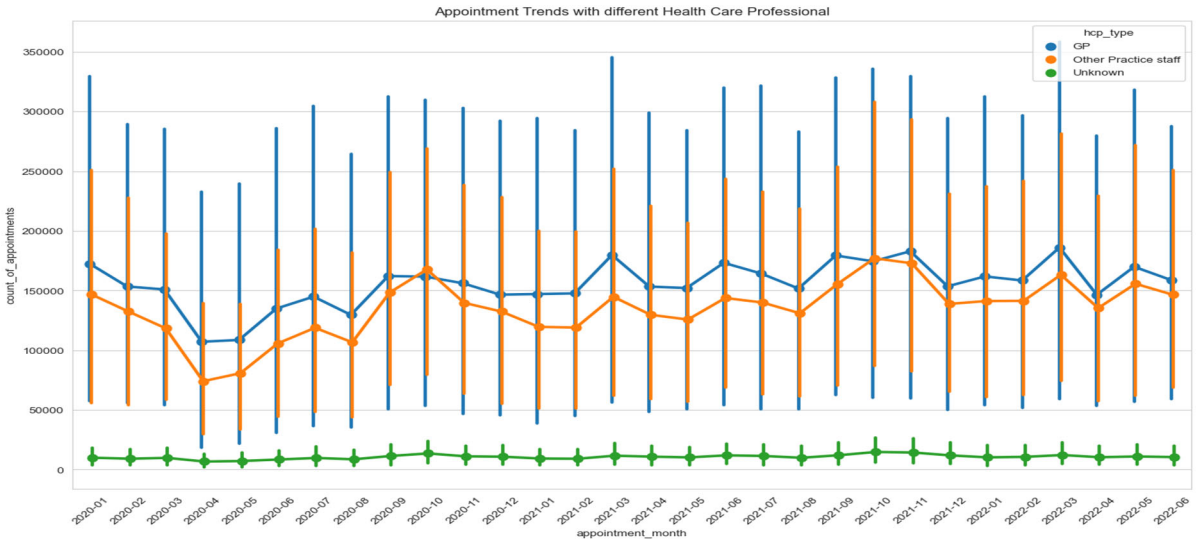
INSIGHTS

The records from 30-month data revealed that the capacity of the utility were under deployed.

Attention should be paid in order to avoid wastage of resources and to develop incentives for the community to fully take advantages of the services offered from NHS.

BUSINESS QUESTION:

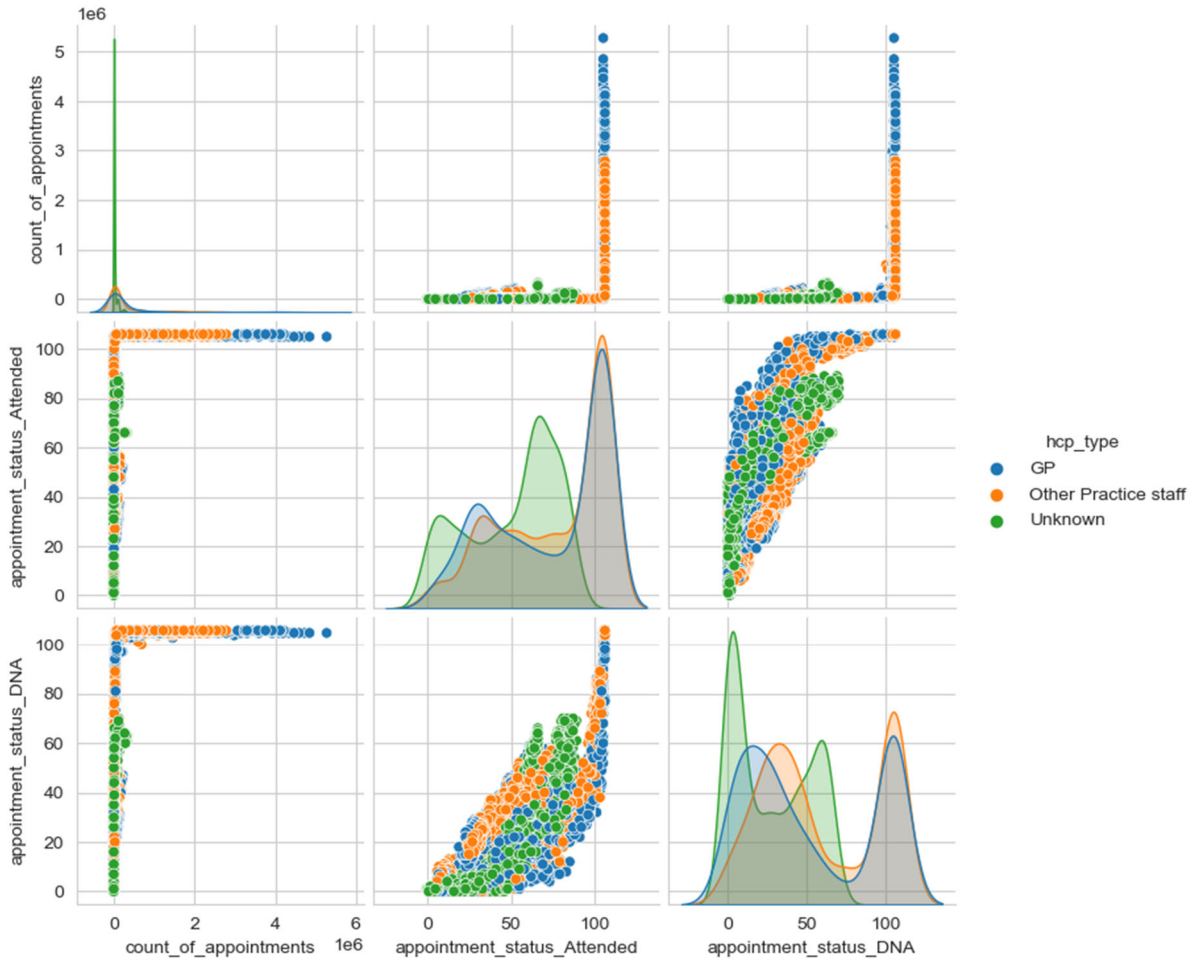
How do the healthcare professional types differ over time?



INSIGHTS FROM THE HCP APPOINTMENT PER MONTH

The trends between GP and Other Practice staff basically showed a positive relationship. However, the spread of the data from GP was obviously larger than that of the Other Practice Staff.

Due to a large number of centers of NHS the availability of other practice staff in each center was not known. Also there might have mis-matching of services to the client, for example some patients with special care for which general practice consultation might not help but required for specialized services.



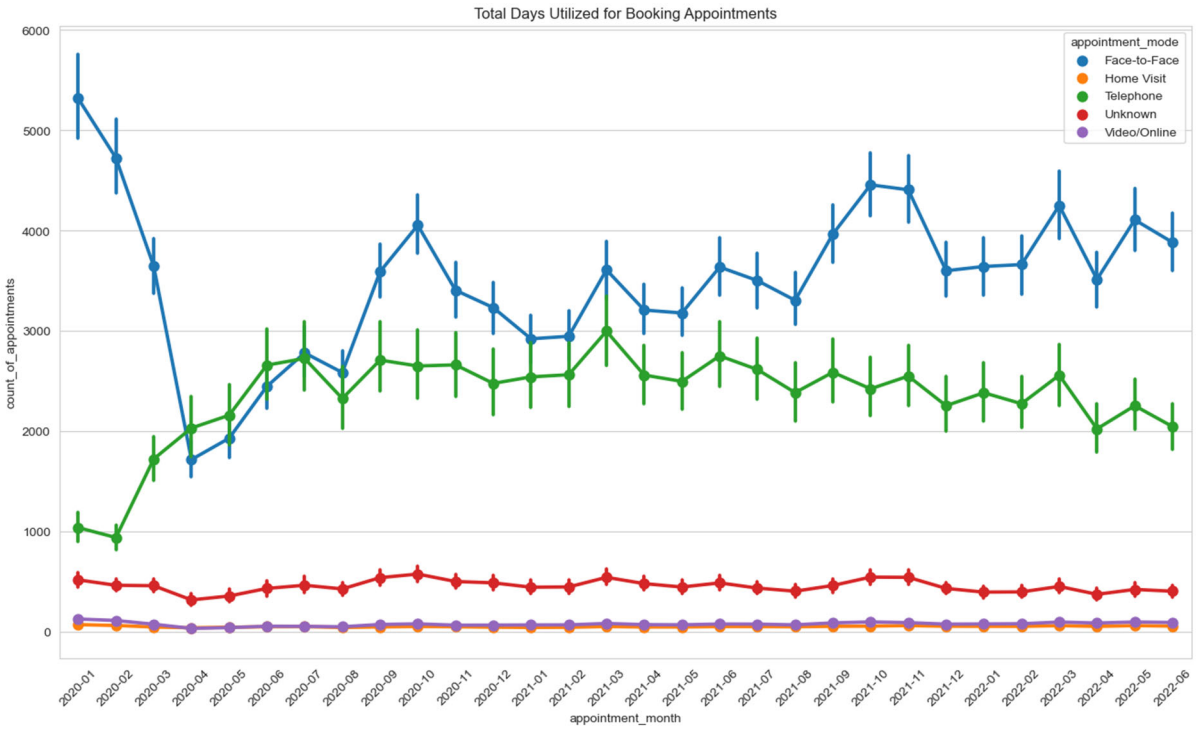
INSIGHTS ABOUT ATTENDANCE STATUS¶

1. From the pairplot of the total attendance and do-not-attendance, the hcp type of might have different pattern to each category.

2. For appointment recorded as attended, the distribution of count (in the middle of the above pair plot), the distributions from each hcp category were some kind a bi-modal distribution, i.e. there existed two highest frequency of counts. It would be obvious that the distribution pattern for the "unknown" hcp were different to the rest of hcp's. The pattern for GP and Other Practice staff showed similar pattern of distribution

3. While for those appointment recorded as do-not-attend (in the lower right corner), the distribution between the three types of practioners were varied between each group. A general observation were that the unknown practioners skewed towards to the of the plot, which might be due to some systematic reasons instead of a probability of chance of attendance comparing to the other practices.

4. The double peaked distribution usually indicates you've got two different groups in the sample



INSIGHTS FROM THE TREND

1. There was a tremendous dropped down of number of appointment starting from Jan 2020 - Apr 2020. This might be attributed to the COVID19 restriction policy announced by the Government in order for limiting the infection of the virus.

Sources: https://www.instituteforgovernment.org.uk/data-visualisation/timeline-coronavirus-lockdowns

2. After the ease of the restriction, the face-to-face appointment climbed steadily and the trend was moving upward in the plot.

3. While for appointment mode with telephone, it appeared that there was a downward trend up to the end of the studied period.

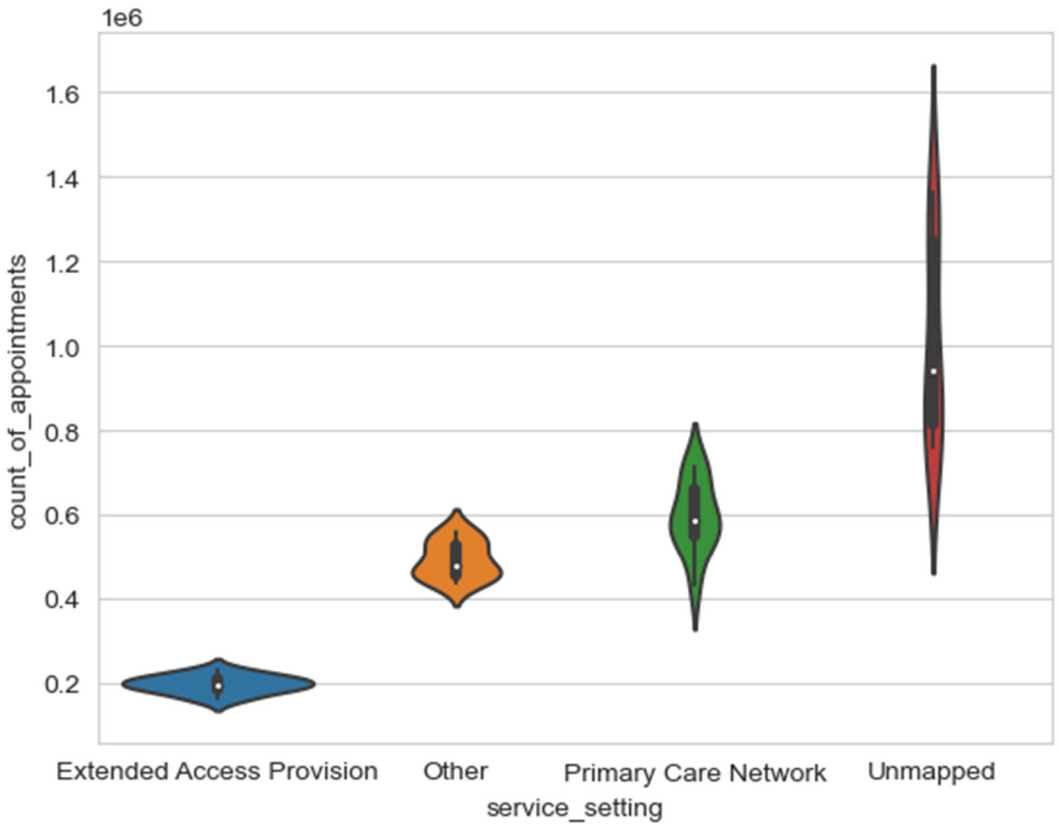
4. Therefore, a plan for moving resources from telephone appointment to face-to-face might be wise decision, although the cost for face-to-face appointment might be higher than the telephones.

5. For the rest of appointment modes, the chart showed a slightly downward trends accordingly.

SUGGESTION:

The trend of different types of appointment mode tended to show a mild downward trends with the count of appointment. The community seemed moving away from the services provided with the current mode of appointments.

Action should be taken to attract the community to utilize the NHS services.



INSIGHT FROM THE SERVICE SETTING WITHOUT GP

1. The Extended Access Provision occured with the least spread and symmetrically distributed with the counts of appointment meaning that this group presented an accurate mean of estimate.

2. The unmapped group was showing in the plot with a very large spread of data which might attribut to some purely random occations.

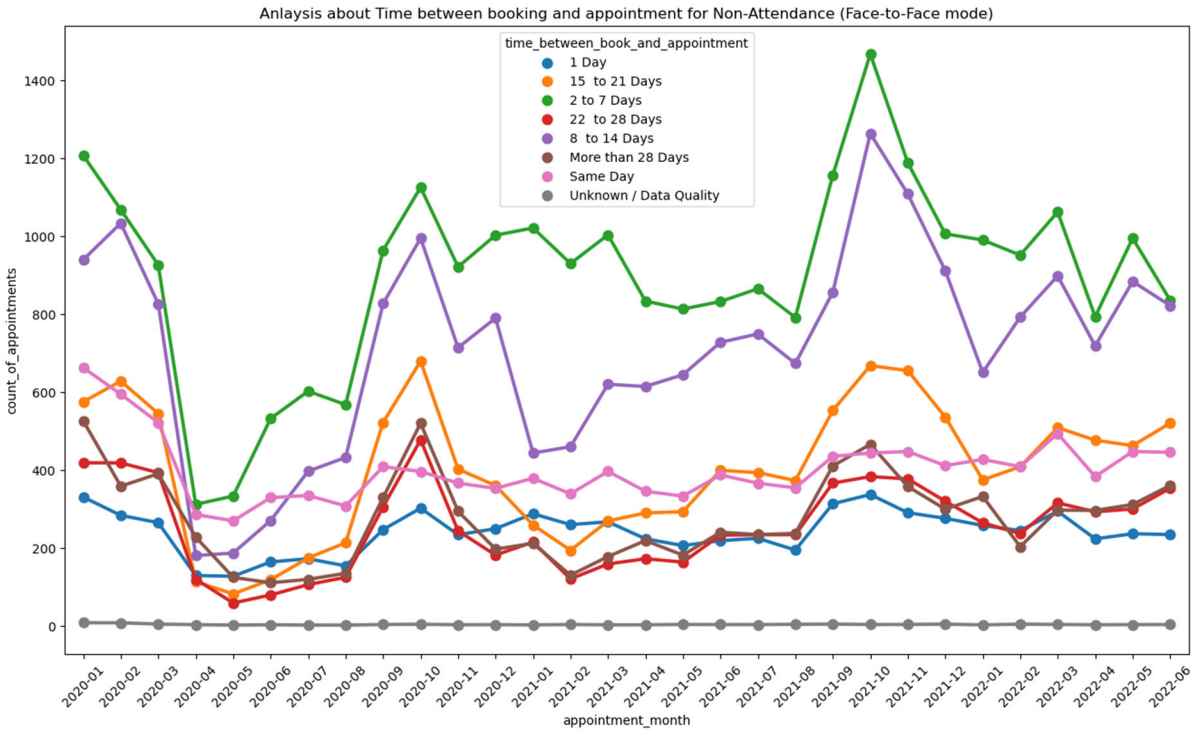
3. For the group primary care network, the data showed a wider spread but still tended to be symmetrically distributed. This might be of less error during input.

4. The "Other" group appeared to be skewed with two mode within the data occurance, which might have some more than one major variable inside.

EXPLORATION OF REASONS FOR NON-ATTENDANCE

While it was shown that the face-to-face mode of appointment contributed to the highest count of appointment, it would be necessary to know further about the reason of non-attendance accordingly.

We would work with the sub-category for all-non-attended class with the mode of appointment as follow.



SUGGESTIONS

The time between booking and appointment for 2-7 days was of the highest count of appointment which were UNATTENDED compared to other categories, and more than 28 days being the second highest.

It was also noticed that the time between booking and appointment on the same day being the third highest in the unattended criteria, which might demonstrate the reason for not attending the appointment might not be attributed to the scheduled time.

The reason for unattendance could be complex and required further analysis, like questionnaire survey conducted during the booking conversation to know further if the patient's preferred schedule could match with the GP appointment schedule.

CONCLUSION

The NHS services were under-utilized on the one hand and there was a downward trend of appointments with NHS professionals. This might be one of a reflection of the substantial awareness from the social media, e,g. Tweeters, in which topics about healthcare were highly awarded.

The reason might not be attributed to the time between book and appointment and further effort is suggested for data about the quality of services, patient history and demographics

Another suggested approach regarding the services arrangements could refer to the study on the seasonal data, where it indicated a strong preference as to appointment schedule planning for better allocating services following the identified patterns.