Background

The National Health Service (NHS) has contracted us to investigate the reasons why patients missed General Practitioner (GP) appointments. Every missed appointment will have a financial cost to the NHS and a social cost to the patient, if we can understand the reasons, then we may be in a position implement a strategy that can reduce the overall volume of missed appointments ensuring a net benefit to the NHS and patients alike.

It is possible that there is not adequate staff to provide the services that patients need? How can we analyse the data to determine if this is the case?

How are the GP resources distributed across the difference services?

Is there a way to assess which appointments will be more likely to not be attended by a patient?

Analytical Approach

The raw data files were large, and importing from CSV/Excel each time into a Pandas DataFrame was not efficient, so PostgreSQL was used to store the original datasets. A database named ""nhs_data" was created, to be used as a central repository for the raw data, the wrangled data and any subsets. Subsequently, stored data can be queried through SQL with ease. Fig. 1 illustrates a visual display of the raw data structure.

Initial analysis of the raw data revealed the following:

1. Locations are arranged by 3-levels: Sub-ICB (106 locations), ICB (42), Region (7). The dataset for appointments_regional was aggregated to the ICB level, so for this report all datasets were subsequently aggregated to the ICB level.
2. Descriptive ICB and Region location text was missing.
3. Appointment_date varied between datasets, most notably appointments_regional was aggregated by month and began in the month of January 2020.
4. Appointment_date/appointment_month column for actual_duration and appointments_regional datasets were the incorrect datatype which was fixed.

The process of preparing the data for analysis began with (i) importing and sense-checking the raw data from PostgreSQL, (ii) data wrangling (fixing datetime datatypes, merging location data, reformatting large strings, reordering columns), (iii) export final dataset to PostgreSQL, (iv) import and sense-check final dataset. Fig. 2 shows the final structure of the data. Using SQL, it will be simple to combine the datasets on the ICB code value.

The location data used was extracted from ONS (Office for National Statistics) Geography portal[1]. Population and GP data related to ICB locations were

downloaded from the NHS Website[2]. The data was combined using Excel and imported into the project database.

A user-defined function written to aggregate a column (e.g., date or context type) by sum of appointments and percentage of the total appointments allowed for efficient analysis of each column, and provided many useful insights as detailed in the Jupyter Notebook. One of these was the impact of Covid-19 which created anomalies in the data (see Fig. 3). Covid status[3] was mapped to each month to allow for effective visualisation. Based on this, further analysis of the appointments_regional dataset was reduced to the "post-Covid" period (August 2021 onwards), which would also bring the data in line with the national_categories dataset.

GitHub version control was implemented to keep track of changes to the code.

Visualisation and Insights

1. National Categories

(a) GP Daily Appointments

- Regular GP Surgeries open Mon-Fri (Fig. 4).
- Tuesday is the busiest day (Fig. 4).

Fig. 5 shows timeline of daily appointments by season. Autumn and Winter (October, November & December 2021) regularly exceeded daily capacity, generally considered to be approximately 1.2 million appointments per day. Summer (August 2022). In October and November 2021, GPs opened on Saturday, presumably planned due to increased patient demand. During August 2022, daily appointments dropped below capacity. December 2021 shows closures during the Christmas holiday period, which explains the dip in appointments shown on other charts.

(b) Context Type

This metric is not useful to the analysis. Fig. 6 shows the "Unmapped" category reducing dramatically over time. The interpretation is that the data will be more reliable if appointments are mapped correctly.

(c) Service Setting

- General Practice is 91.5% of appointments (Fig. 7).
- Timeline analysis illustrates that The Primary Care Network has seen increased demand, and General Practice usage peaked during October 2021 (Fig. 8). 2 Month SMA smoothed the output line.

(d) National Categories

- Total of 18 national categories. The top category is General Consultation Routine (Planned, Follow-up, Review) (Fig. 9).
- Top 7 categories account for 96.33% of appointments (Fig. 9).

Increasing demand for General Consultation Acute (Emergency, NHS 111, Same Day). Planned Clinics & Clinical Procedure increased suddenly and peaked in October 2021, then reduced quickly, but remains higher than in August 2021 (Fig. 10).

2. Actual Duration

53.13% of appointments last less than 15 minutes (Fig. 11). The high level of Unknown / Data Quality category suggests the data could be unreliable. The fact the data is not mapped to any particular category means the analysis is limited to location and date.

3. Appointments Regional

(a) Healthcare Professional Type (Fig. 12)

- 50.12% of appointments by GP.
- High demand in October 2021 and March 2022.
- Other Practice Staff not sub-categorised.

(b) Mode (Fig. 13)

- 61.98% of appointments were Face-to-Face.
- 33.87% of appointments were by Telephone.
- Appointments by Telephone gradually decreasing over time.

(c) Time Between Book & Appointment

- 73.5% of appointments seen within 1 week of booking (Fig. 14).
- Appointments more than 2 weeks after booking increasing towards October 2021 levels (Fig. 15).

(d) Appointment Status

- Average of 4.5% of appointments are missed.
- London and North West record the highest percentage of missed appointments (Fig. 16).
- Daily missed appointments are below average since December 2021 (Fig. 17).

Patterns and Predictions

Analysis of missed appointments by region illustrates:

1. London has low numbers of GPs per 100000 population and high % of missed appointments (Fig. 18).
2. North West high numbers of GPs per 100000 population and high % of missed appointments (Fig. 18).
3. The other regions show a positive correlation between % of missed appointments and the sum of appointments (Fig. 19).
4. The trend is the longer patients wait between booking the appointment and the actual appointment, increases the chances of that appointment being missed (Fig. 21).
5. London is a hotspot for missed appointments (Fig. 22).

Appointment waiting times greater than 1 week can double the probability that the appointment will be missed. Therefore, we can say that efforts should be made to reduce waiting times between booking and appointment date.

Increased demand for acute (emergency, same-day) services may negatively affect planned services (Fig. 10).

Key recommendations:

- Healthcare Professional (HCP) types: the data needs to be improved, to include more detail about which type in the "Other Practice Staff" category.
- Actual Duration: the data will be more useful if it can be assigned to a national category and/or HCP type, then we can see how long appointments take relating to the task.
- National category could include the data for missed appointments.
- Target improving the time between booking and appointment. Some appointments will require the longer time gap, but for others aiming for a shorter gap (to less than 1 week) may reduce the number of missed appointments.
- Anonymised social media sentiment analysis can be utilised during the busiest periods to identify when staff shortages are affecting patient satisfaction.

Further analysis should be focused on:

- The usage of resources by national category; compare the ICB locations with high/low number of missed appointments to identify any trends or relationships.
- A more granular look into the Sub-ICB location level, in particular the North West region, may show how the usage of resources differ.
- GPs per 100000 population are low in the London region, could there be other staff shortages (e.g., Nurse, Physio) which increase the length of time between booking and appointment date?

References

1. Sub ICB Locations to Integrated Care Boards to NHS England (Region) (July 2022) Lookup in England: https://geoportal.statistics.gov.uk/datasets/2bca16d4f8e4426d80137213fce90bbd_0/explore

2. NHS Website (GP Data & Population by ICB Location): https://www.england.nhs.uk/wp-content/uploads/2022/04/integrated-care-board-allocation-core-services.xlsx

3. UK Parliament Covid Data: https://commonslibrary.parliament.uk/research-briefings/cbp-9068/

Appendix

Fig. 1: ERD (Entity Relationship Diagram)

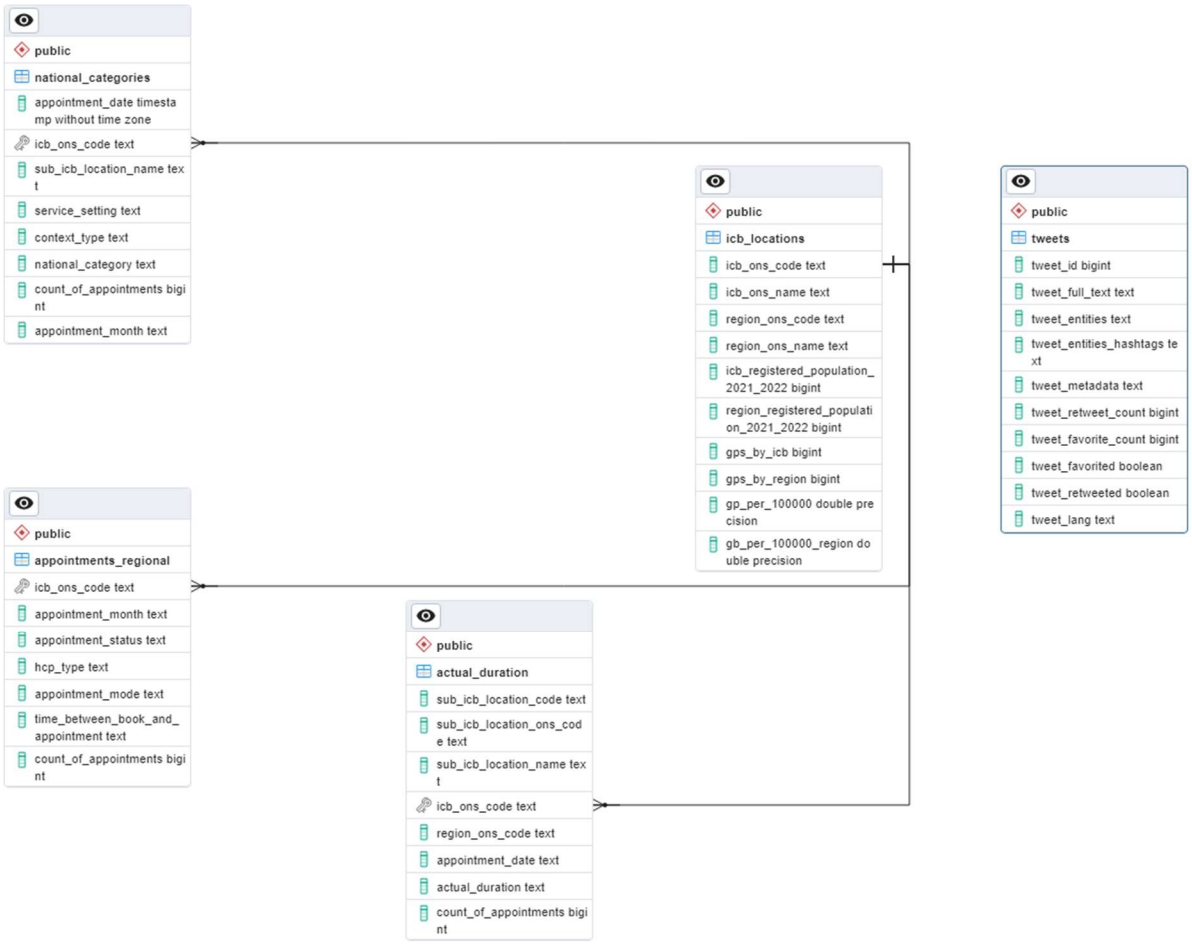


Fig. 2: ERD Final

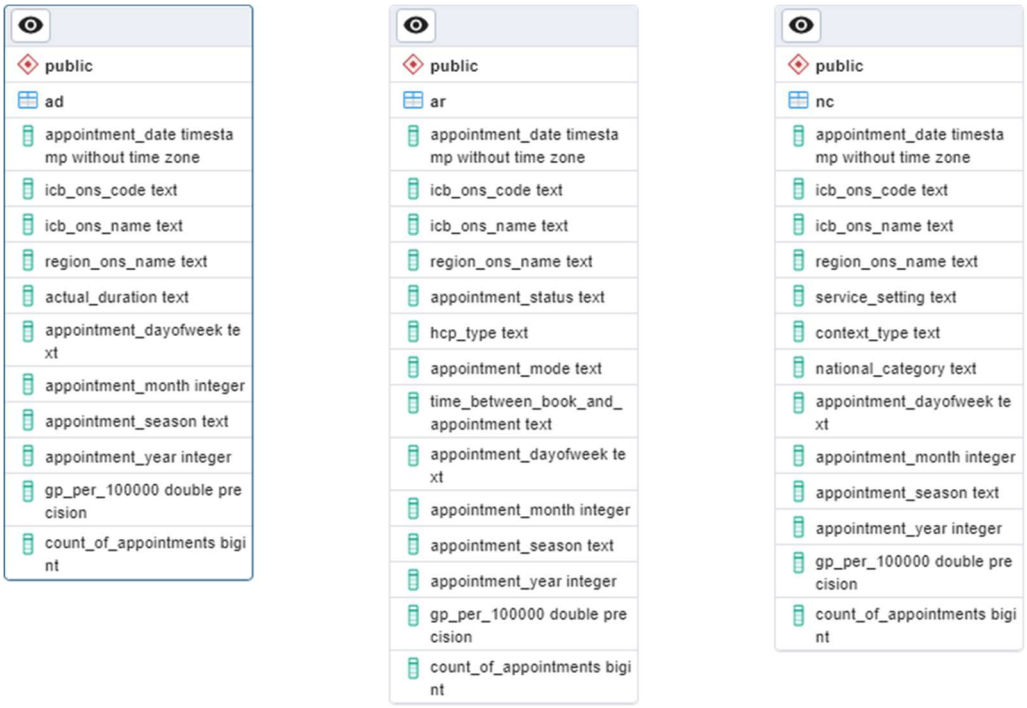


Fig. 3: Covid Timeline Analysis

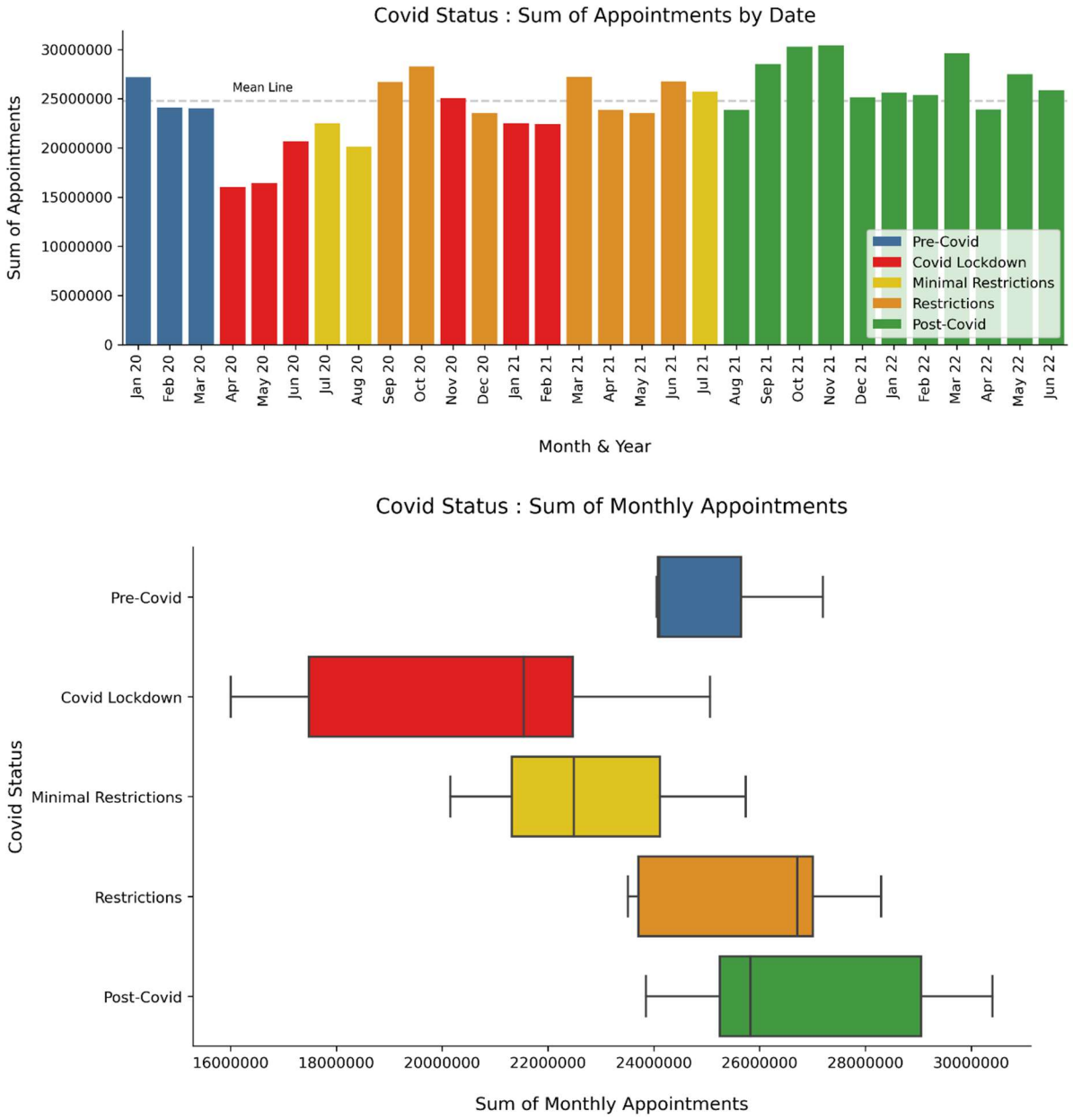


Fig. 4: General Practice Appointments by Day of the Week

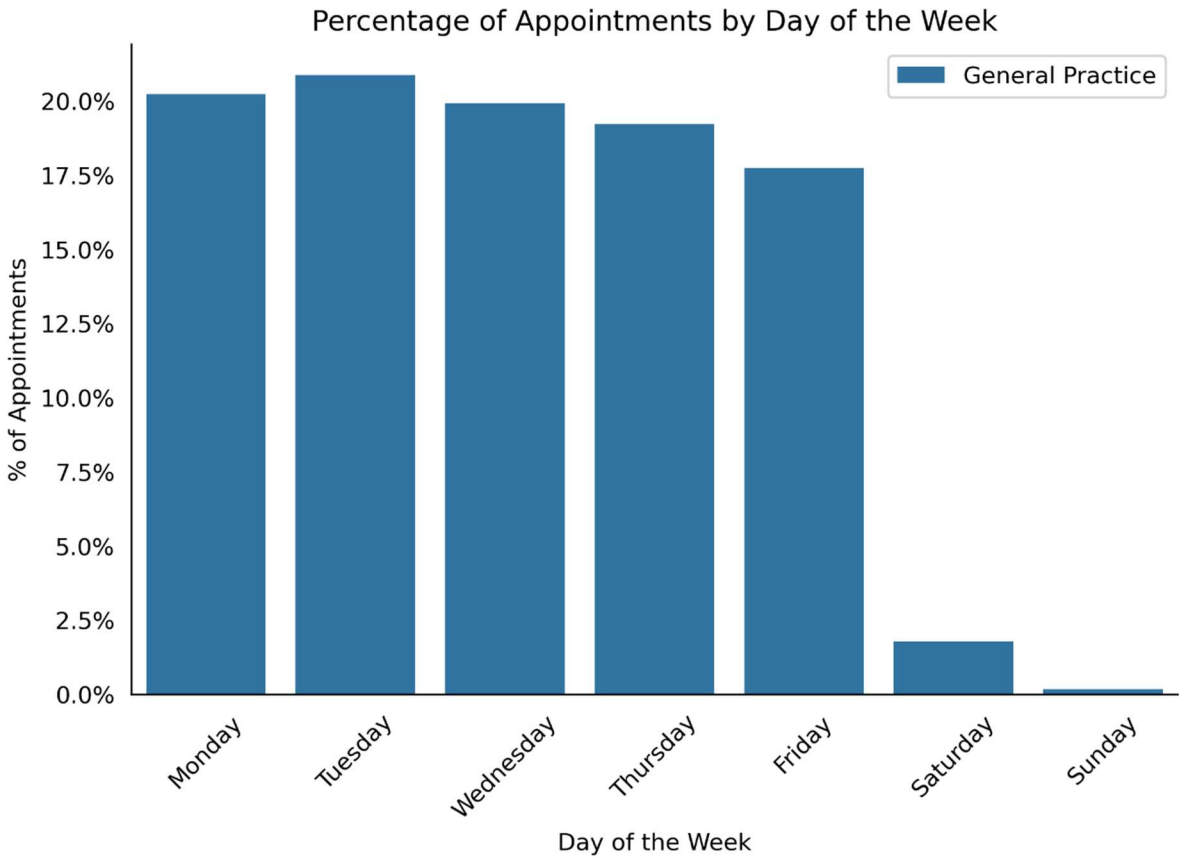
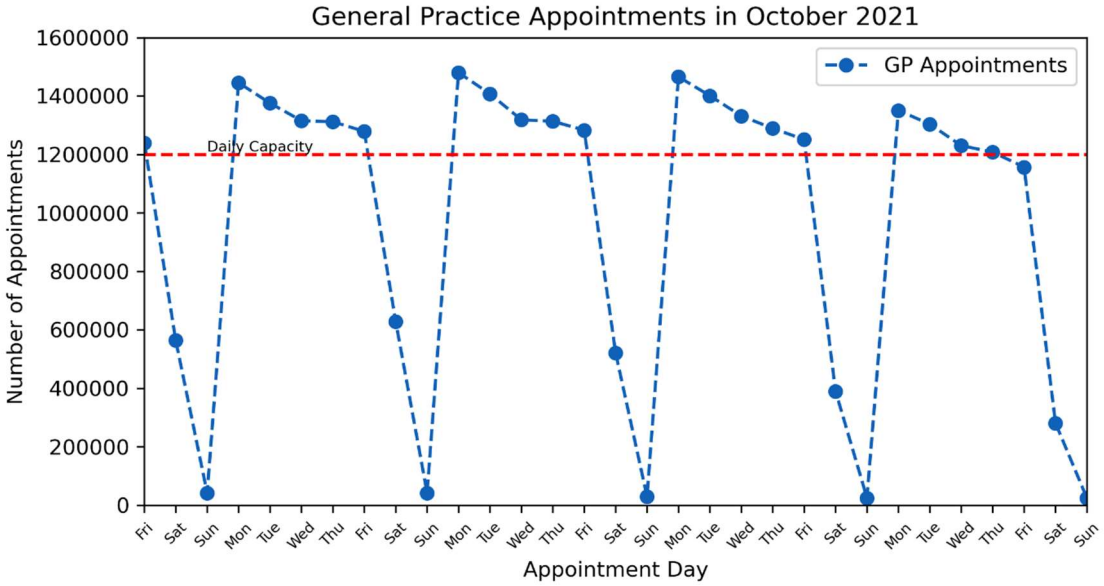
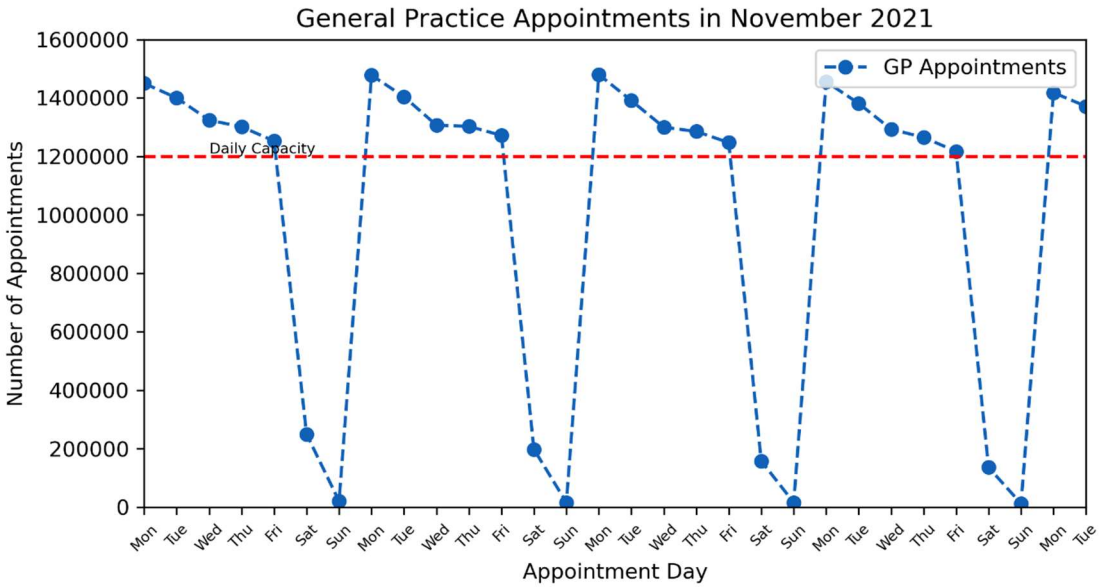
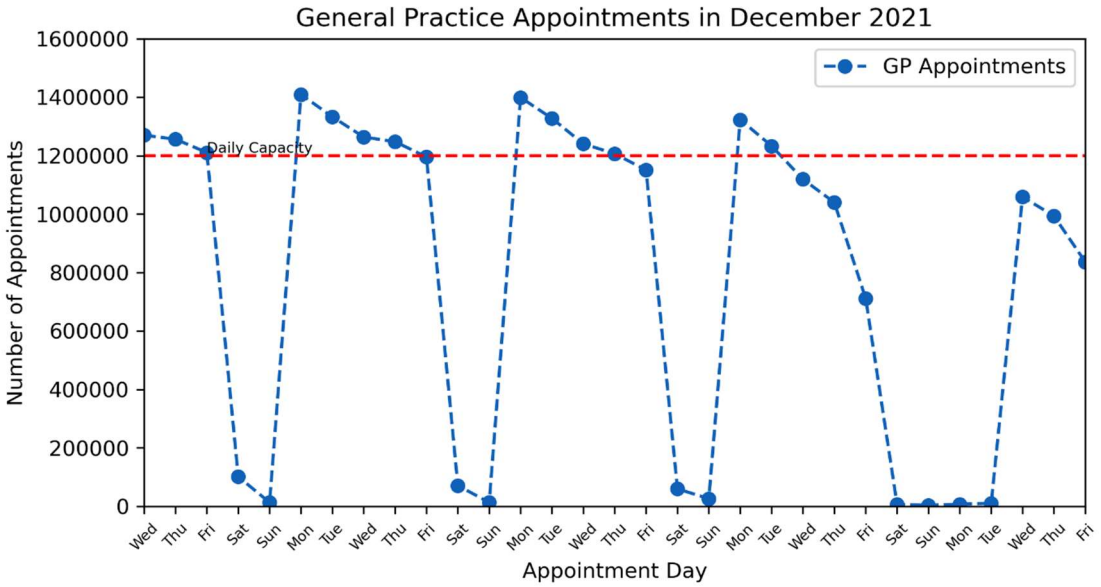


Fig. 5: Daily General Practice Appointments by Season




General Practice Appointments in November 2021
GP Appointments
Daily Capacity
Number of Appointments
Appointment Day


General Practice Appointments in December 2021
GP Appointments
Daily Capacity
Number of Appointments
Appointment Day

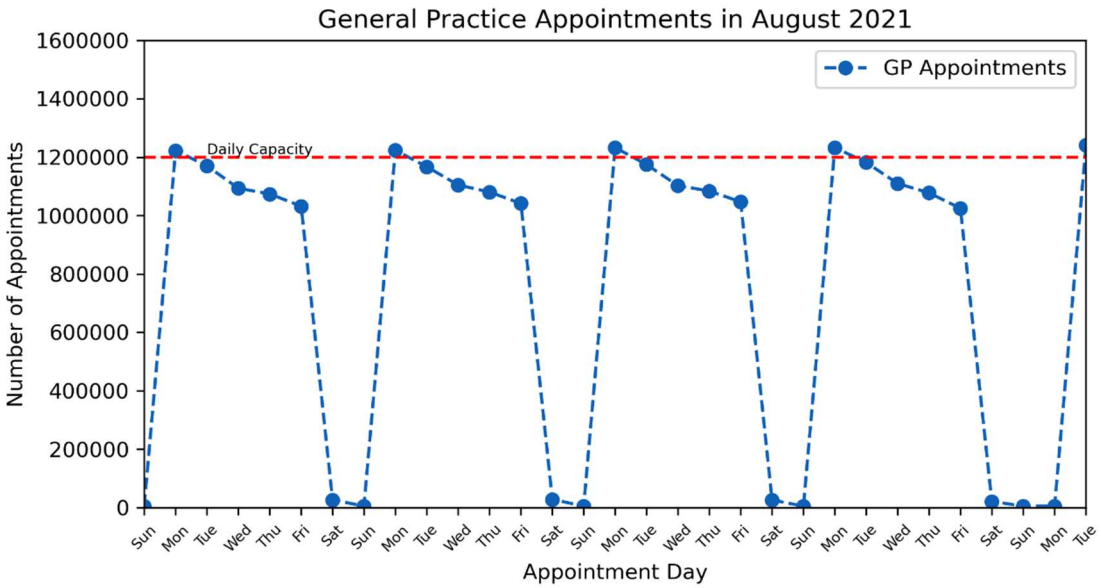

General Practice Appointments in August 2021
GP Appointments
Daily Capacity
Number of Appointments
Appointment Day

Fig. 6: Appointments by Context Type: Timeline Analysis

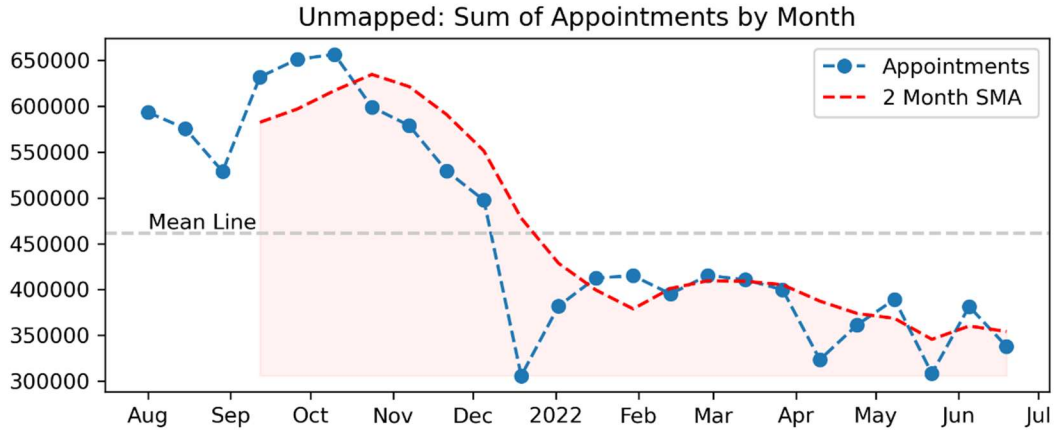


Fig. 7: Appointments by Service Setting

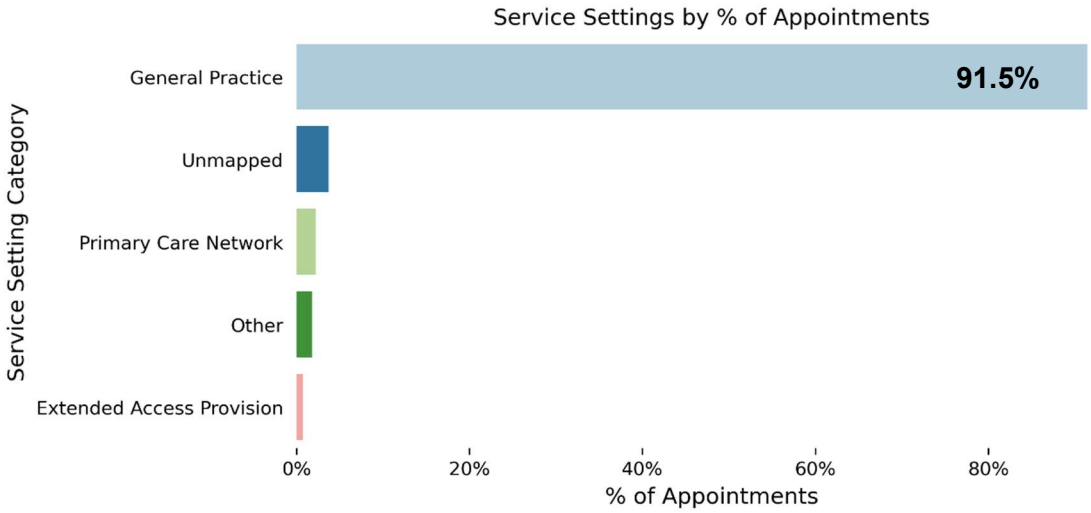


Fig. 8: Appointments by Service Setting: Timeline Analysis

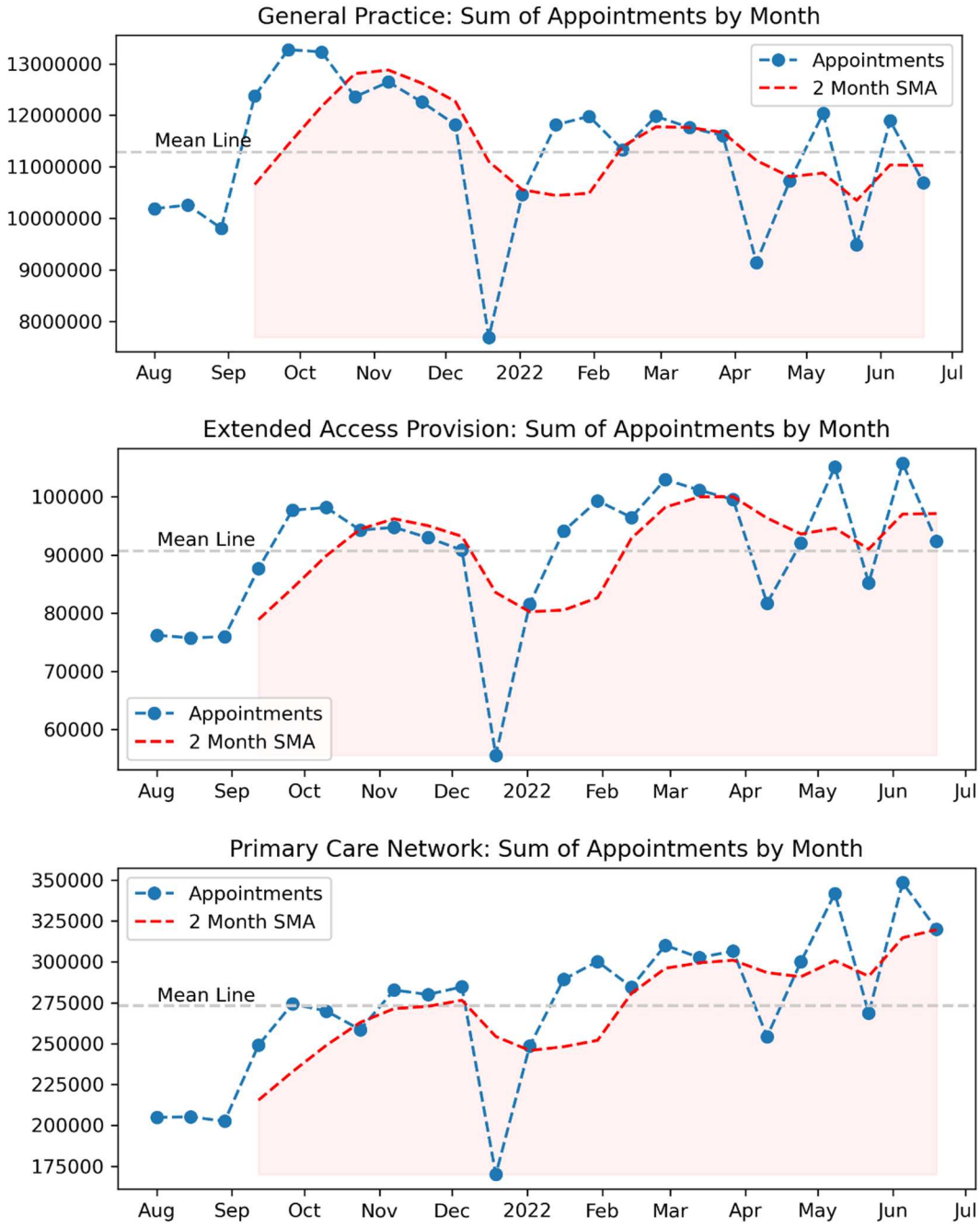


Fig. 9: GP Appointments by National Categories

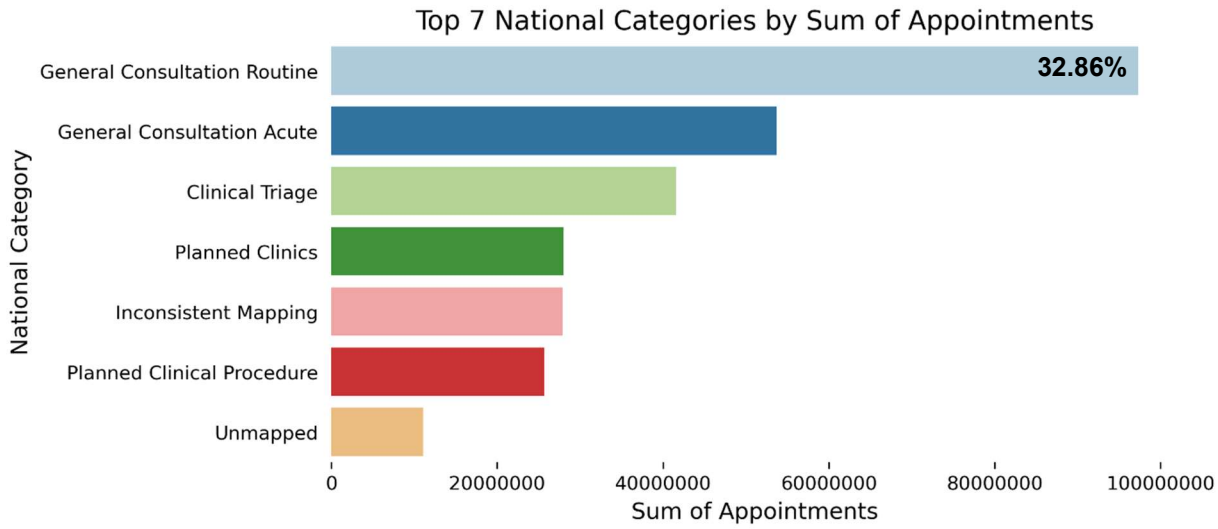


Fig. 10: Monthly GP Appointments by National Categories

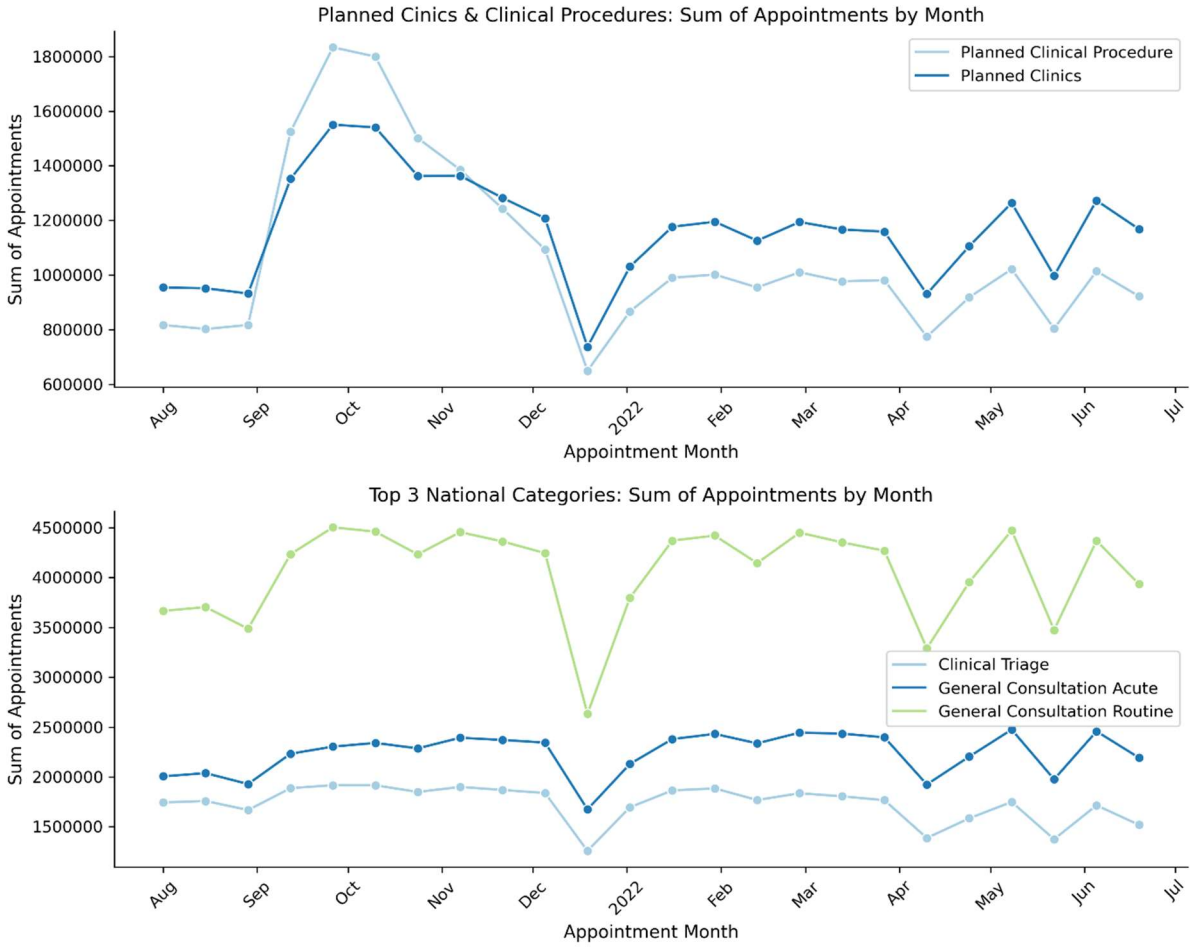


Fig. 11: GP Appointments by Actual Duration

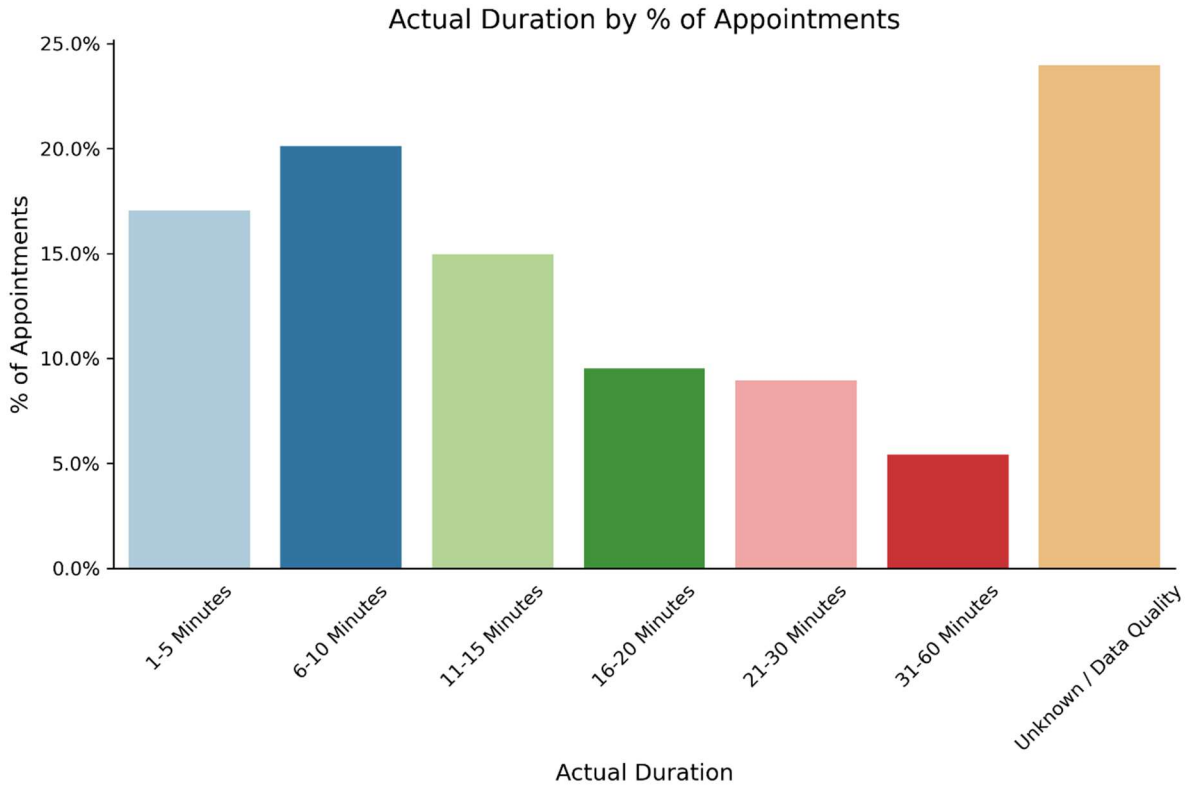


Fig. 12: GP Appointments by Healthcare Professional

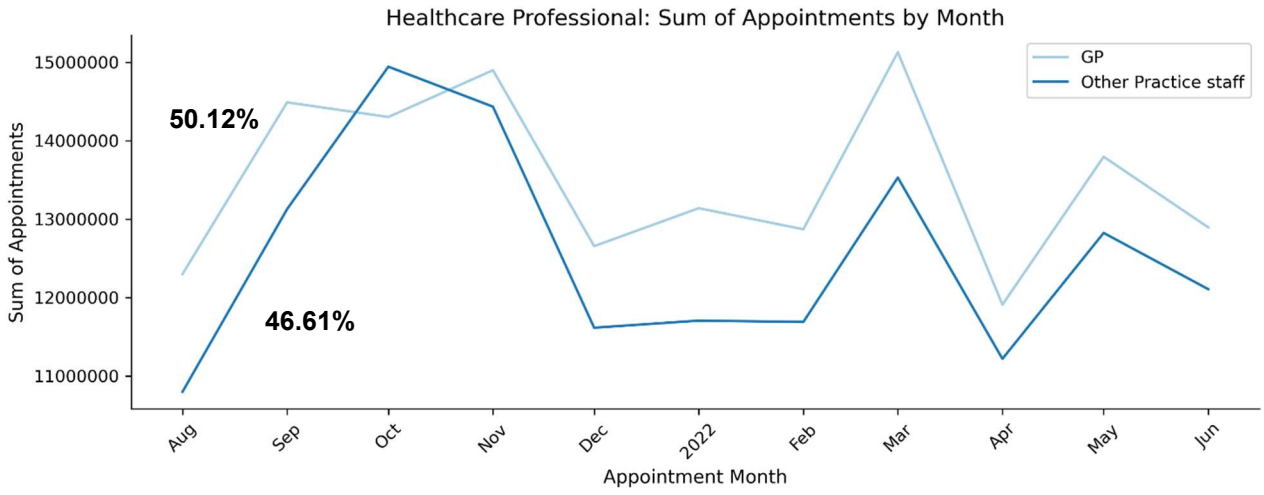


Fig. 13: GP Appointments by Mode

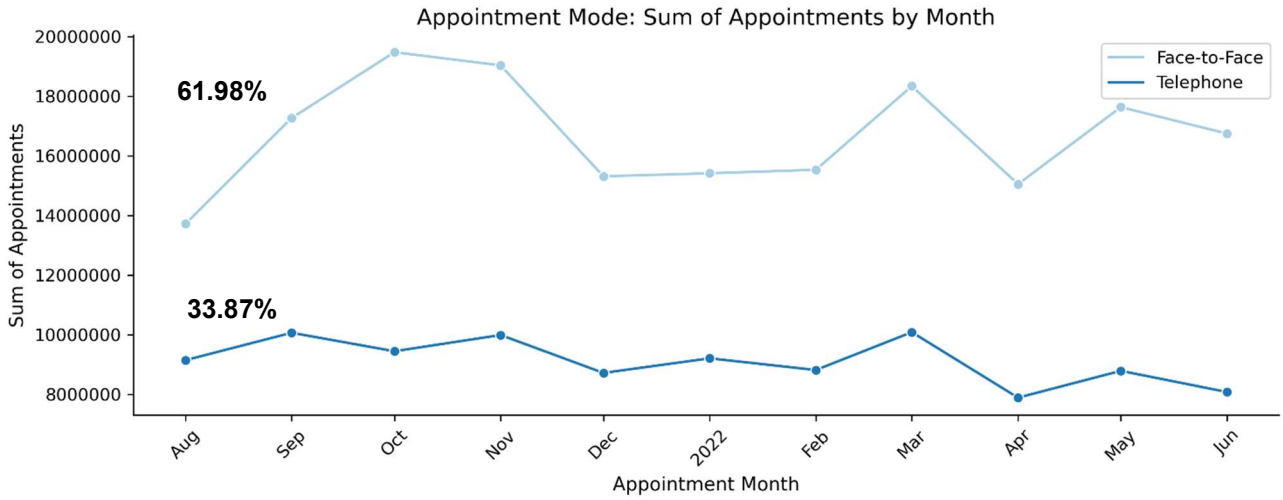


Fig. 14: GP Appointments by Time Between Book & Appointment

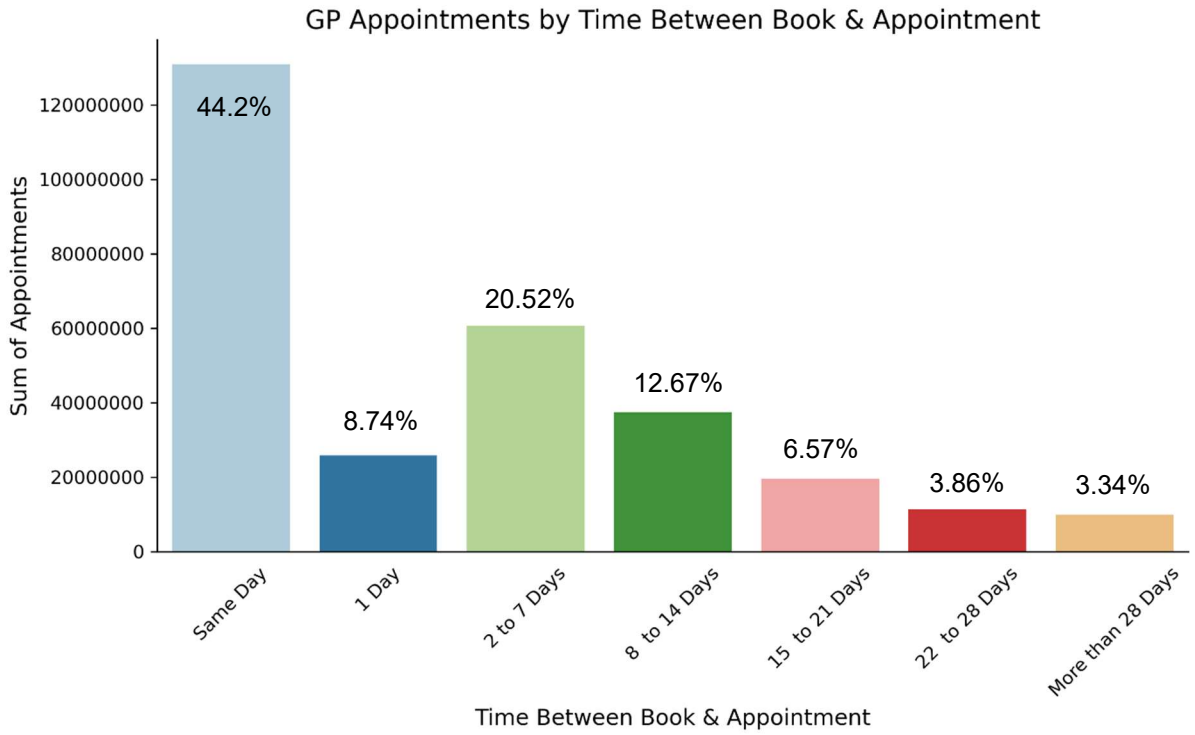


Fig. 15: GP Appointments by Time Between Book & Appointment

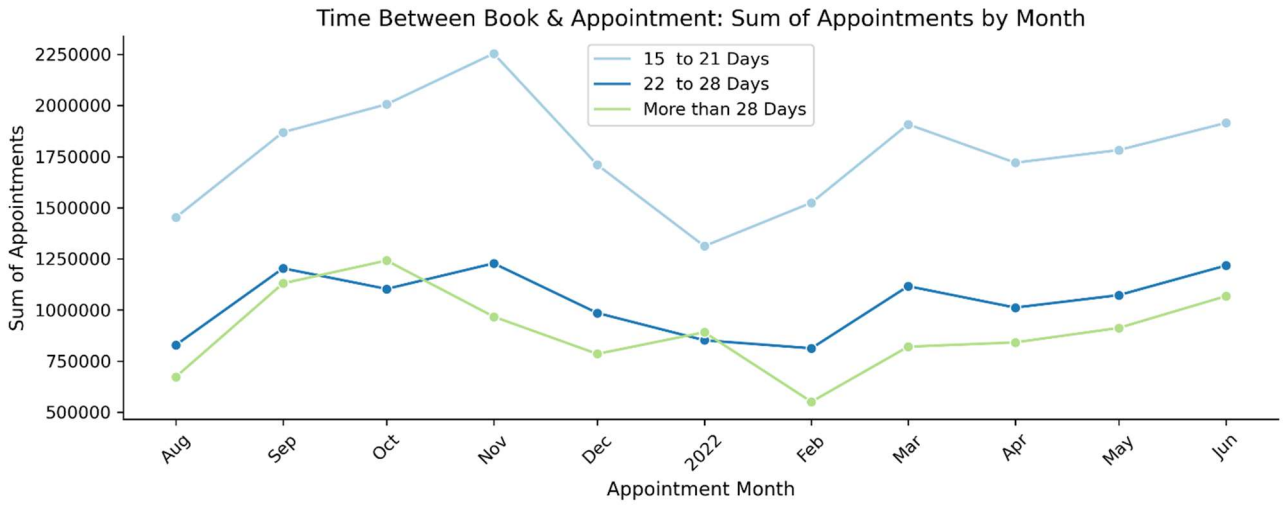
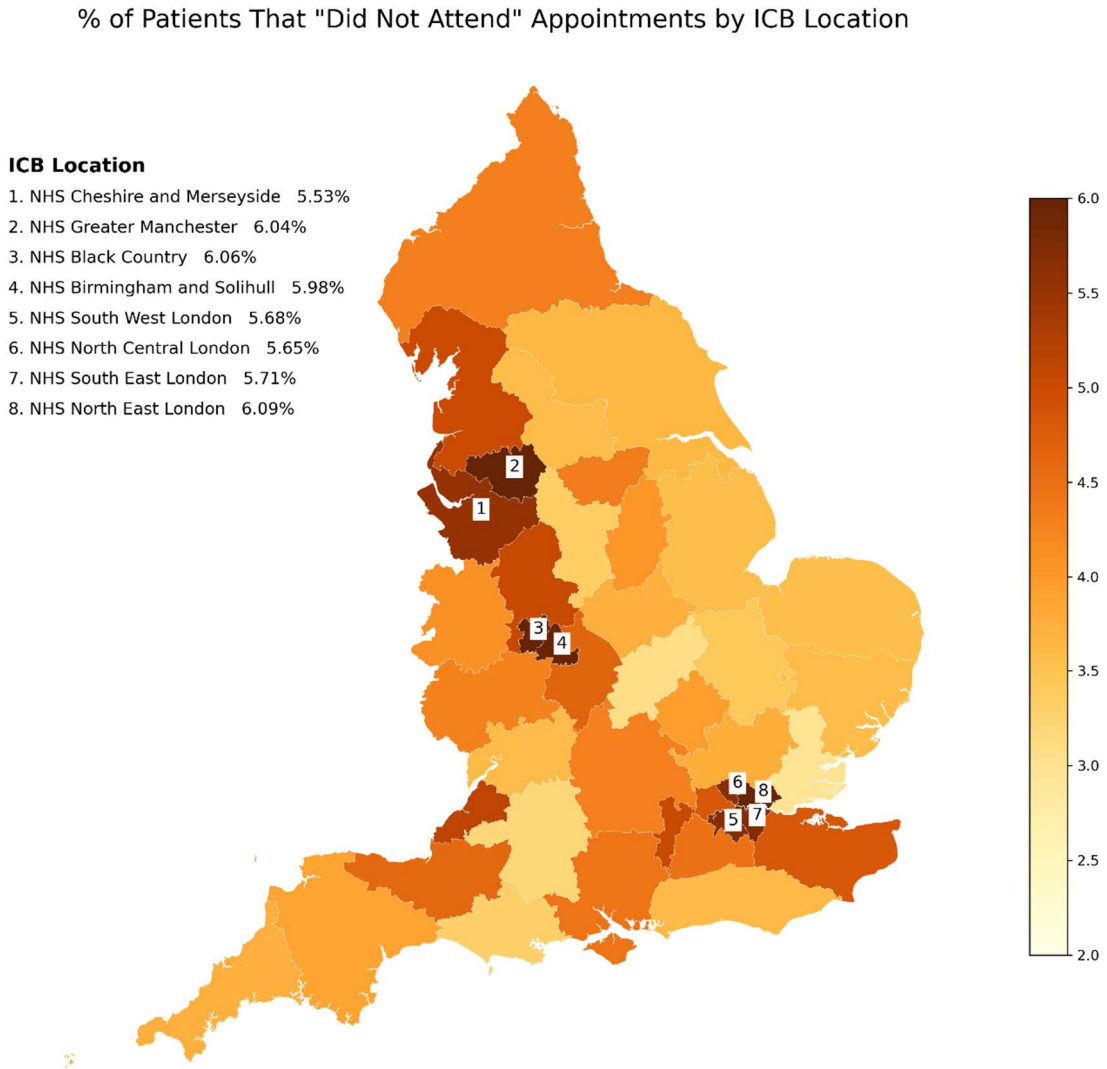


Fig. 16: GP Missed Appointments by ICB Location



Source: Government Statistics 08/2021 to 06/2022

Fig. 17: GP Missed Appointments by Month

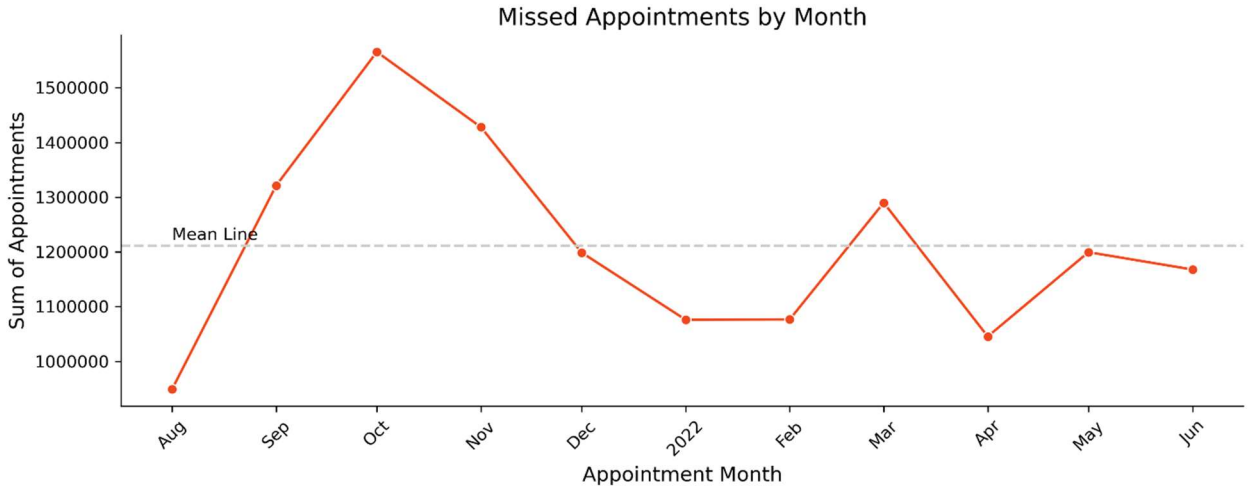


Fig. 18: % and Sum of Missed Appointments by ICB Region & GP per 100000 Population

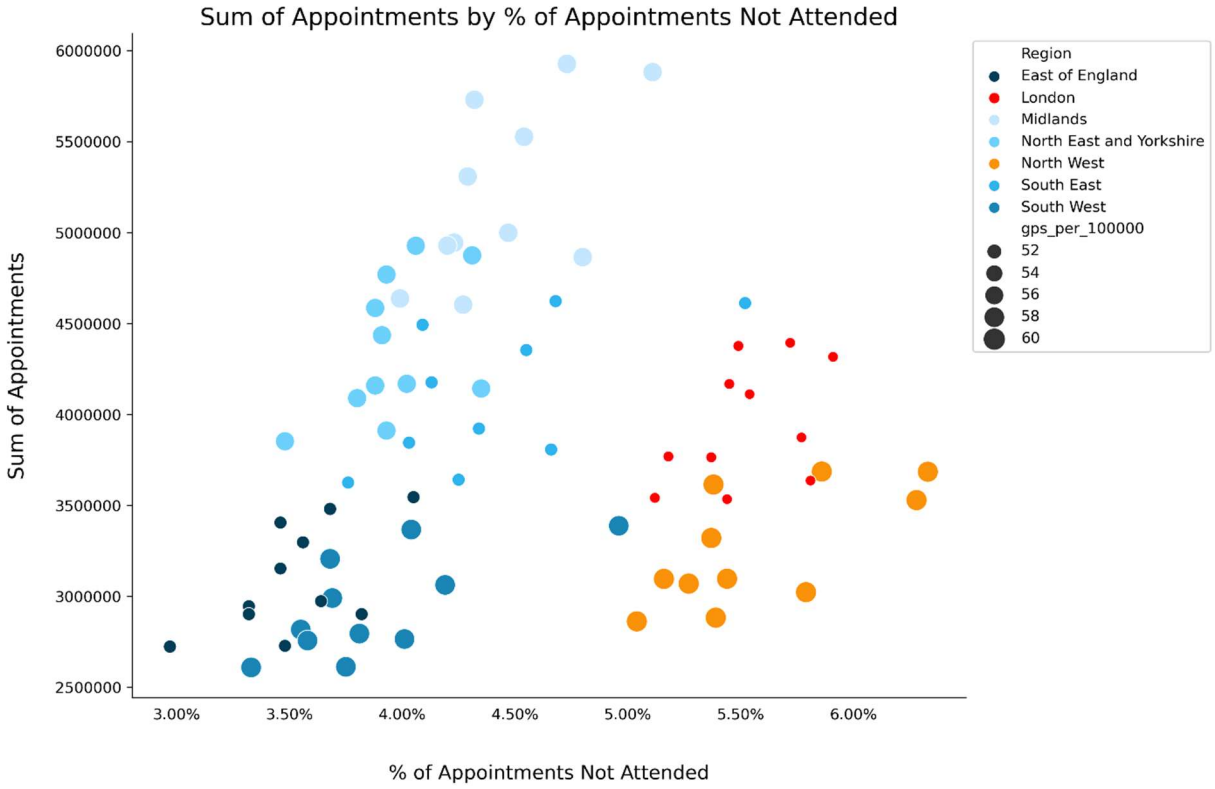


Fig. 19: % and Sum of Missed Appointments by ICB Region (Excluding North West & London)

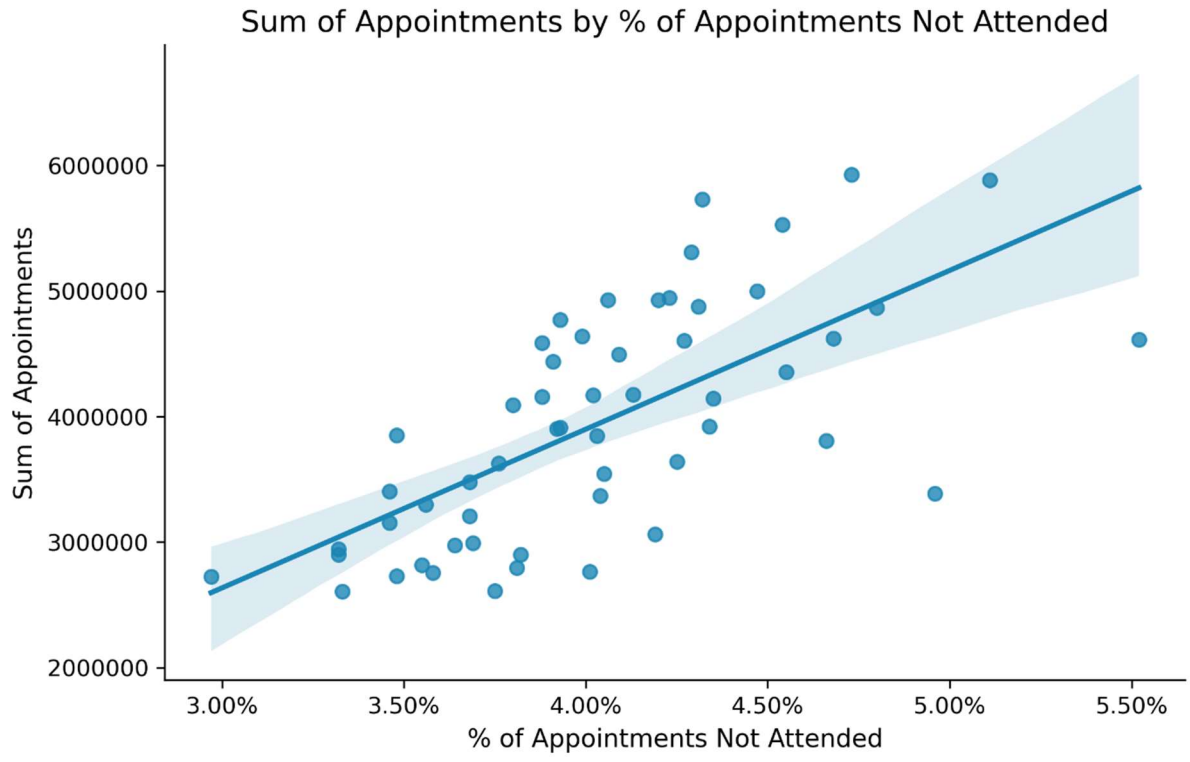


Fig. 20: % and Sum of Missed Appointments by Region: Timeline Analysis

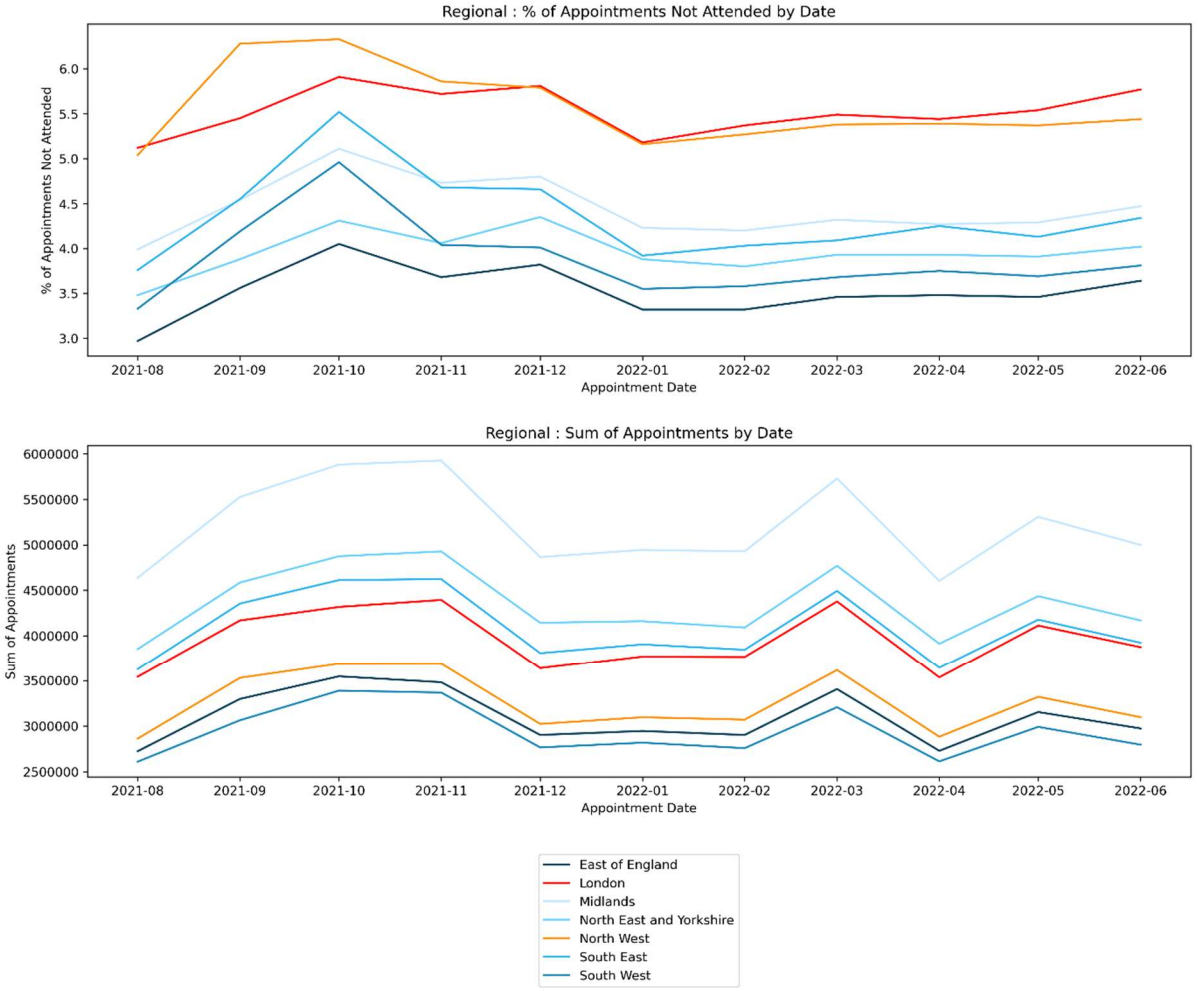


Fig. 21: Missed Appointments by Time Between Book & Appointment: All Regions

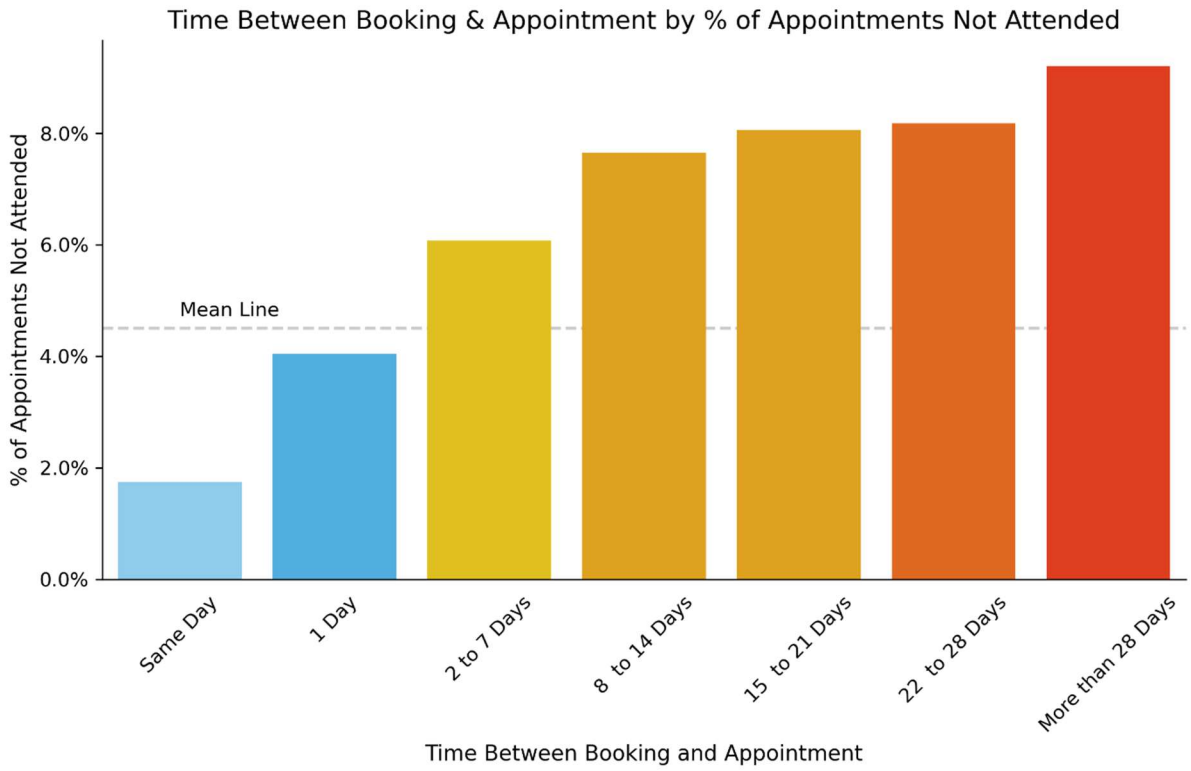


Fig. 22: Missed Appointments by Time Between Book & Appointment: All Regions vs. London