Introduction

The NHS faces rising demand for healthcare services, leading to concerns about staff capacity, appointment efficiency, and service accessibility. Our analysis identifies key trends in appointment attendance, scheduling efficiency, and workforce needs, providing data-driven recommendations for improved resource allocation.

This analysis examines the need for staff expansion, changes in healthcare roles over time, shifts in appointment attendance rates, seasonal variations in appointment types and volume, the impact of scheduling time on attendance, and differences across service settings.

Data Cleaning & Handling

No null values have been detected as the data was cleaned before it was provided.

Duplicates were found in the appointments_regional dataset, accounting for less than 0.1% of total appointments. These were removed to ensure integrity. Investigating the root cause of duplication is recommended to prevent recurrence.

Python was chosen due to the large dataset size and the need for in-depth analysis and visualization. Other tools would require combining several, making Python the most practical choice.

Data Source and Quality

The three datasets—**national_categories (nc), appointments_regional (ar), and actual_duration (ad)**—show inconsistencies in appointment counts for the same months:

- **nc** reports the highest number of appointments.
- **ar** is slightly lower but still relatively close.
- **ad dataset** contains the fewest records.

These discrepancies are attributed to reporting limitations, as indicated in the metadata. To prevent inaccuracies when analysing the various segments, the datasets were not merged.

To ensure transparency, data handling decisions prioritized accuracy over artificial adjustments:

- Unknown/unmapped values were retained to avoid distorting an accurate picture. For example, if 91% of appointments were attended, 5% missed, and 4% unknown status, removing the unknown values would inflate the attendance rate to 95%.
- Seasonality analysis is limited due to the nc dataset covering only August 2021 – June 2022.
- Staff role lack of classification occurs, with only GP roles accurately recorded, while others are grouped under "Other Practice Staff."

- A deeper analysis at the Sub-ICB level is needed to provide more granular regional insights.

Findings

Total Locations: 106

Top 5 Locations by Appointment Count (Aug 2021 - June 2022)

```
sub_icb_location_name
NHS North West London ICB - W2U3Z              6976986
NHS North East London ICB - A3A8R              5341883
NHS Kent and Medway ICB - 91Q                  5209641
NHS Hampshire and Isle Of Wight ICB - D9Y0V    4712737
NHS South East London ICB - 72Q                4360079
Name: count_of_appointments, dtype: int64
```

General Practice accounts for 91.5% of all appointment service settings, emphasizing its role in primary healthcare (Appendix 1).

Care related encounter is the only context type provided, suggesting that additional types may be used but are not listed (Appendix 2).

Categories General Consultation Routine (32.9%) and General Consultation Acute (18.1%) highlight their importance in preventative and urgent care (Appendix 3).

91.3% of appointments attended. Analyzing non-attendance trends could help identify potential causes for non-attendance (Appendix 4).

Analysis and Trends

Appointment Data Range:

- nc: Aug 2021 - June 2022
- ad: Dec 2021 - June 2022
- ar: Jan 2020 - June 2022

(Appendix 5)

Busiest Month: November 2021 across all 3 datasets. Over 30 million appointments across all locations. This is most likely related not only to flu season which happens during the colder months, but to rising covid cases and vaccine distribution (source here).

Peak Seasons:

- Autumn: Busiest
- Spring: Second busiest
- Winter: Third busiest

Summer could not be analysed due to data limitations in nc dataset.

Appointment Types & Attendance Trends

Appointments varied by service setting. Non-GP services showed stable volumes, while General Practice fluctuated in line with total appointment trends (Appendix 6).

Care-related encounters followed overall monthly trends (Appendix 7).

Planned clinical procedures peaked in October 2021, while other categories followed general trends (Appendix 8).

Each season is displayed in a bar chart, one bar per service setting, comparing the monthly appointments with an average per month as trendline for autumn, winter and spring data. (Appendices 9–14).

Twitter Data Analysis

Tweet count alone is insufficient for measuring hashtag popularity, as retweets and favorites vary widely (Appendix 15).

Three bar plots compare hashtag popularity by tweets, favorites, and a combined measure. Hashtags appearing in all three categories are highlighted in orange to determine common values. See Appendix 16.

Focus should be on hashtags with the highest visibility, especially those with the most tweets and retweets. See Appendix 17.

Additionally, an attempt was made to identify trending topics by filtering the most frequently used words in tweets. However, the analysis was inconclusive. See Appendix 18.

Potential use cases for this data include sentiment analysis to gauge public opinion on NHS services and trend detection to monitor shifts in healthcare discussions. Challenges remain, such as missing timestamp data, which limits trend analysis, and a lack of resources for sentiment analysis.

Trend Analysis (Aug 2021 - June 2022)

Analysis focused on data from August 2021 for relevance to future decision-making.

Face-to-face appointments accounted for 62%, while 33.9% were conducted by phone (Appendix 19). Among these, 41% of face-to-face and 70% of phone consultations were handled by GP staff (Appendix 20).

Non-GP staff outnumbered GPs only in October 2021. This may relate to a rise in face-to-face appointments (Appendix 21).

Same-day scheduling was most common (44.3%), while only 3.3% were booked more than 28 days in advance (Appendix 22). Evaluating whether faster scheduling aligns with service goals could guide optimization.

Home visits and phone consultations had over 50% same-day bookings, while face-to-face appointments had 35.9% (Appendix 23).

Daily utilization stayed below 1M, peaking at 1.1M in October–November but remaining under NHS's 1.2M capacity (Appendix 24).

Based on National Statistics data (source here), population growth rates were:

- 2019–2020: 0.17%
- 2020–2021: 0.41%
- 2021–2022: 0.99%

Assuming this upward trend continues, 2022 is projected at 1.33%. This was factored into future capacity assessments (Appendix 25).

Appointment status trends were stable (Appendix 26). A shorter booking-to-appointment gap correlated with higher attendance, suggesting shorter scheduling and reminders for long-term bookings may reduce non-attendance (Appendix 27).

Phone and video consultations had the highest attendance. A deeper review of service types could identify more appointments suitable for phone consultations to improve attendance (Appendix 28).

Face-to-face appointments increased in October, aligning with higher non-GP staffing needs (Appendix 29).

The booking-to-appointment time remained stable (Appendix 30).

An ICB-level analysis found no significant regional differences in scheduling patterns (Appendix 31).

Key Recommendations

Increasing non-GP staff availability during peak periods can help manage demand, while expanding phone and video consultations could improve efficiency.

With a 1.2M appointment capacity per month, this limit has not been reached. Further forecasting using the bc dataset is recommended, but staffing increases are not advised.

Same-day appointments show the highest attendance rates, but further analysis is needed to determine if all require immediate scheduling. Identifying which could be managed within a 2–7 day window would optimize resource use. Implementing and testing automated reminders in different formats could help assess their impact.

Unclassified data and inconsistencies in entry reduce report reliability. Improved staff training on accurate data entry would enhance quality. National data entry standards should be implemented for uniformity.

Due to the high cost of acquiring Twitter data, lack of sentiment analysis expertise, and missing timestamps, the value of social media insights remains uncertain. The benefits do not justify the investment.

Appendix 1

	service_setting	count_of_appointments	percentage
1	General Practice	270811691	91.5
4	Unmapped	11080810	3.7
3	Primary Care Network	6557386	2.2
2	Other	5420076	1.8
0	Extended Access Provision	2176807	0.7

Appendix 2

	context_type	count_of_appointments	percentage
0	Care Related Encounter	257075158	86.8
1	Inconsistent Mapping	27890802	9.4
2	Unmapped	11080810	3.7

Appendix 3

	national_category	count_of_appointments	percentage
4	General Consultation Routine	97271522	32.9
3	General Consultation Acute	53691150	18.1
2	Clinical Triage	41546964	14.0
11	Planned Clinics	28019748	9.5
7	Inconsistent Mapping	27890802	9.4
10	Planned Clinical Procedure	25702694	8.7
15	Unmapped	11080810	3.7
16	Unplanned Clinical Activity	3055794	1.0
6	Home Visit	2144452	0.7
14	Structured Medication Review	1858379	0.6
12	Service provided by organisation external to t...	852133	0.3
9	Patient contact during Care Home Round	810330	0.3
1	Care Home Visit	628279	0.2
13	Social Prescribing Service	475828	0.2
17	Walk-in	412438	0.1
0	Care Home Needs Assessment & Personalised Care...	405904	0.1
8	Non-contractual chargeable work	138911	0.0
5	Group Consultation and Group Education	60632	0.0

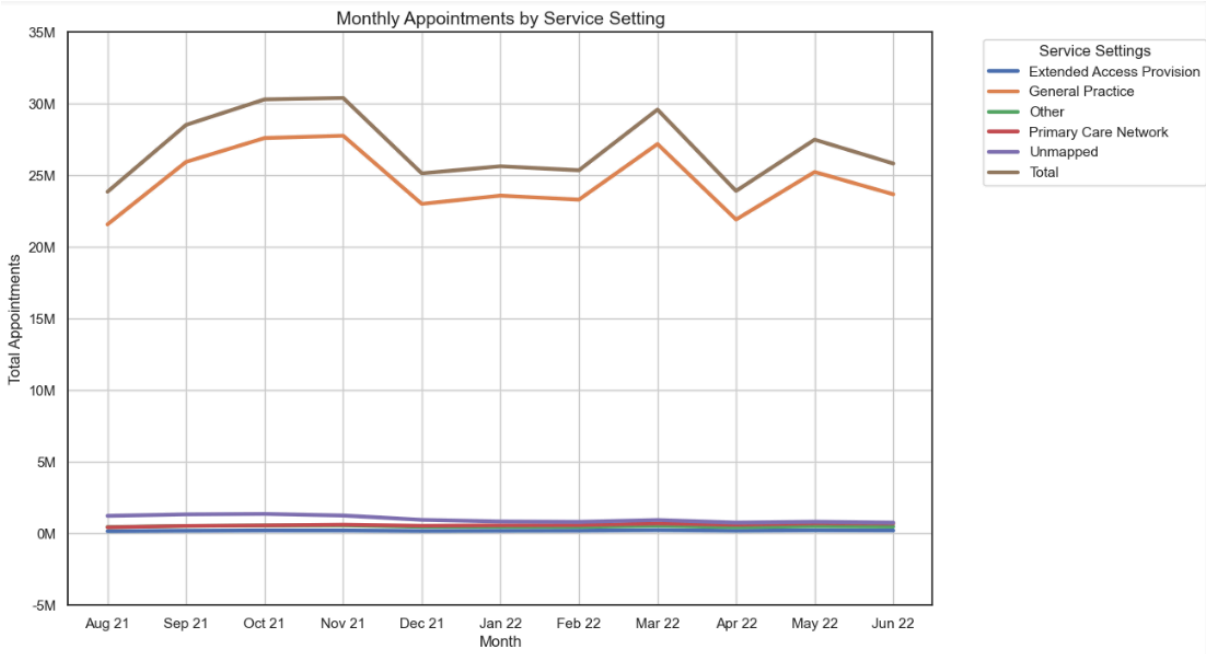
Appendix 4

	appointment_status	count_of_appointments	percentage
0	Attended	677646088	91.3
2	Unknown	34050656	4.6
1	DNA	30833015	4.2

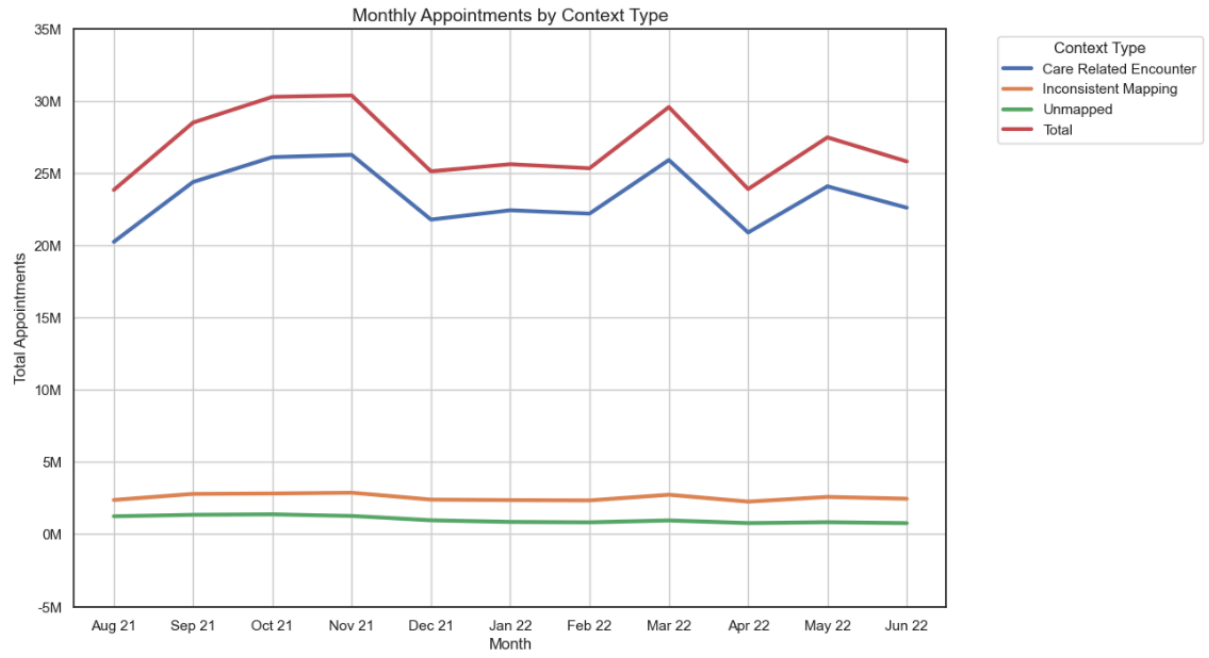
Appendix 5

	appointment_year	appointment_month	ad_appointments	ar_appointments	nc_appointments
0	2020	January	NaN	27191758	NaN
1	2020	February	NaN	24097066	NaN
2	2020	March	NaN	24039635	NaN
3	2020	April	NaN	16002488	NaN
4	2020	May	NaN	16410591	NaN
5	2020	June	NaN	20678416	NaN
6	2020	July	NaN	22483207	NaN
7	2020	August	NaN	20143878	NaN
8	2020	September	NaN	26706462	NaN
9	2020	October	NaN	28296051	NaN
10	2020	November	NaN	25056131	NaN
11	2020	December	NaN	23528063	NaN
12	2021	January	NaN	22482990	NaN
13	2021	February	NaN	22389461	NaN
14	2021	March	NaN	27216974	NaN
15	2021	April	NaN	23872938	NaN
16	2021	May	NaN	23488728	NaN
17	2021	June	NaN	26775653	NaN
18	2021	July	NaN	25728662	NaN
19	2021	August	NaN	23843177	23852171.0
20	2021	September	NaN	28514685	28522501.0
21	2021	October	NaN	30296850	30303834.0
22	2021	November	NaN	30395923	30405070.0
23	2021	December	22853483.0	25132174	25140776.0
24	2022	January	23597196.0	25623928	25635474.0
25	2022	February	23351939.0	25344812	25355260.0
26	2022	March	27170002.0	29586020	29595038.0
27	2022	April	21948814.0	23904960	23913060.0
28	2022	May	25343941.0	27478652	27495508.0
29	2022	June	23715317.0	25819426	25828078.0

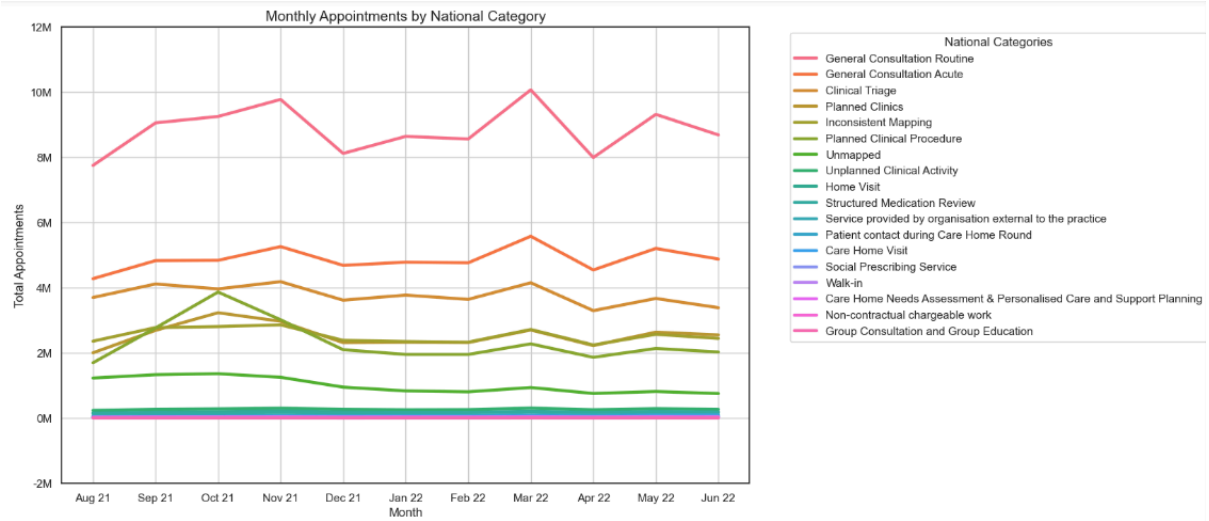
Appendix 6



Appendix 7



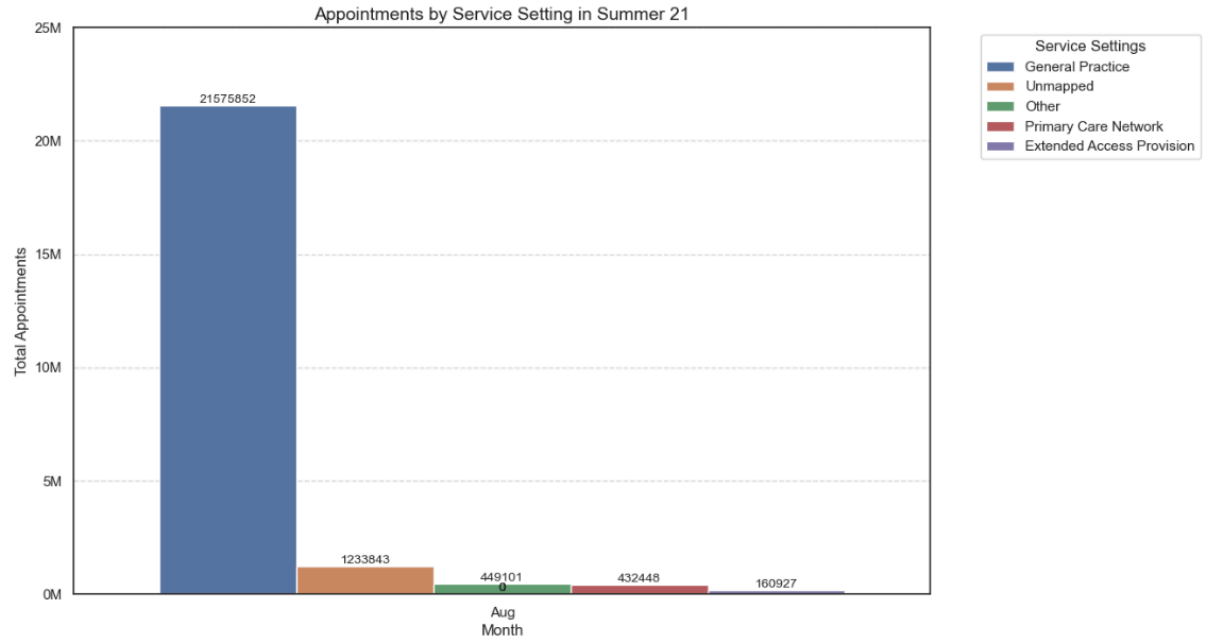
Appendix 8



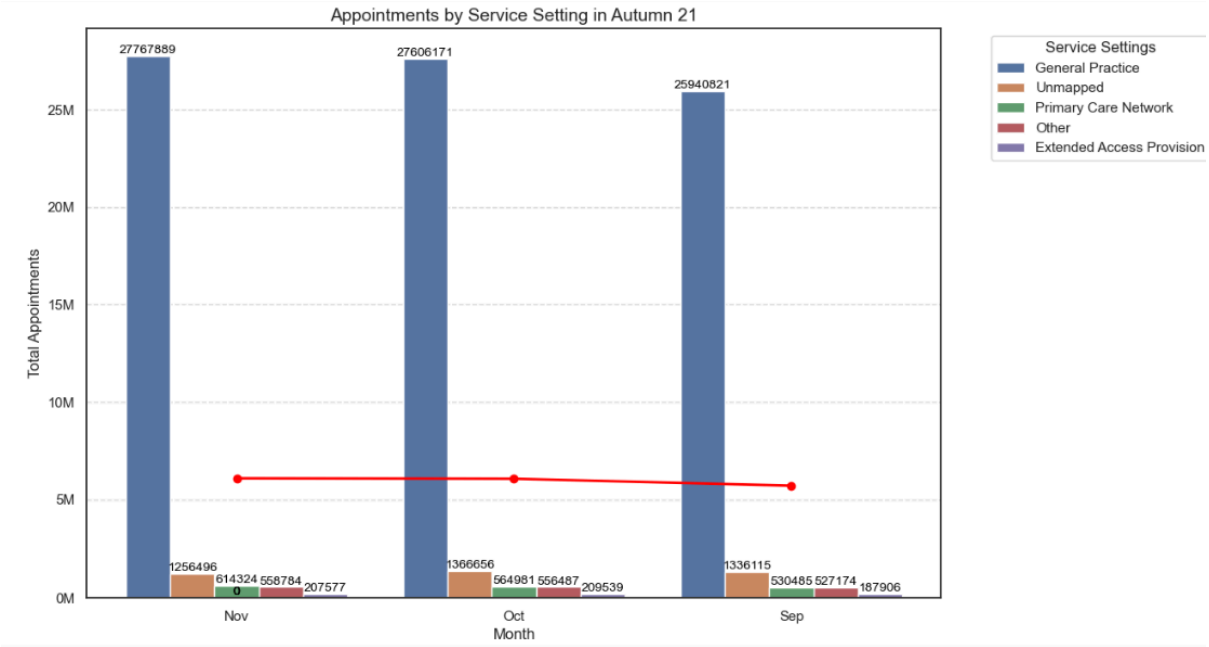
Appendix 9

season	Summer 21	Autumn 21	Winter 21-22	Spring 22	Summer 22
service_setting					
Extended Access Provision	160927	605022	556506	644700	209652
General Practice	21575852	81314881	69897805	74342779	23680374
Other	449101	1642445	1378311	1471406	478813
Primary Care Network	432448	1709790	1693823	2020726	700599
Unmapped	1233843	3959267	2605065	2523995	758640

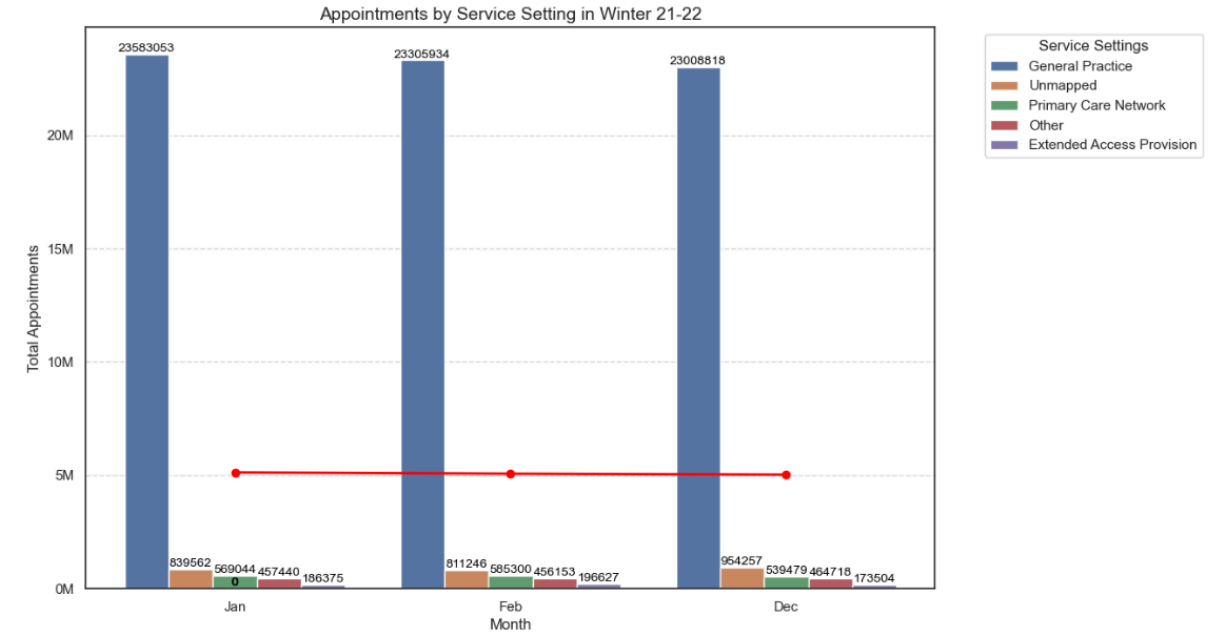
Appendix 10



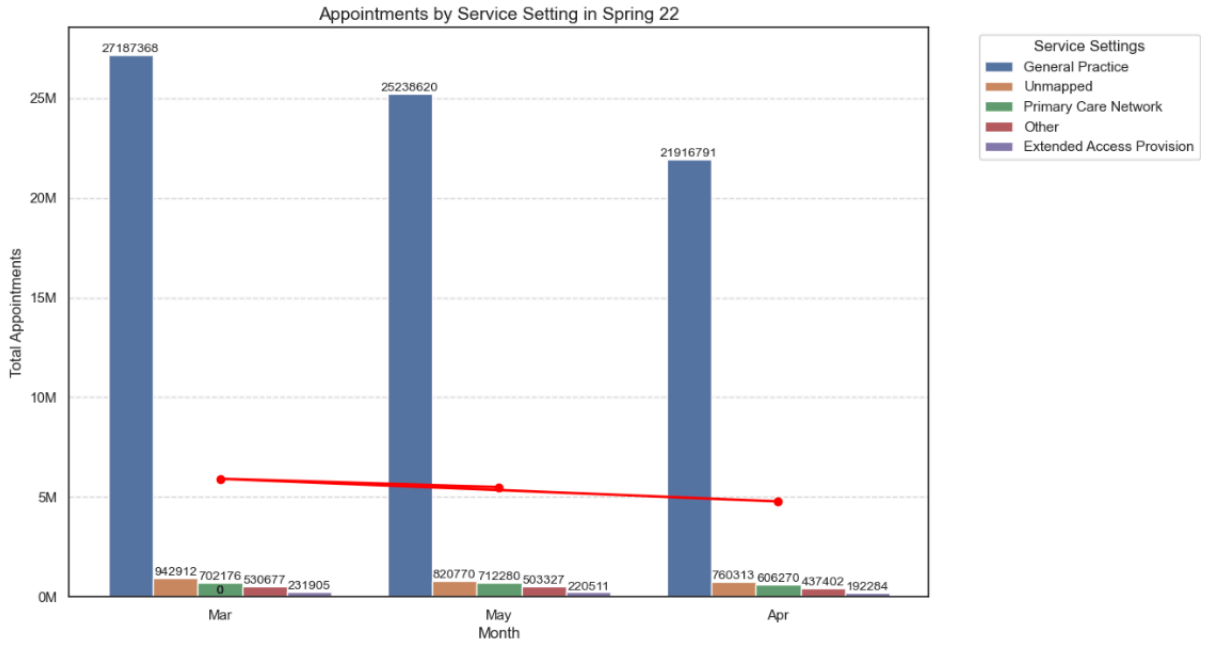
Appendix 11



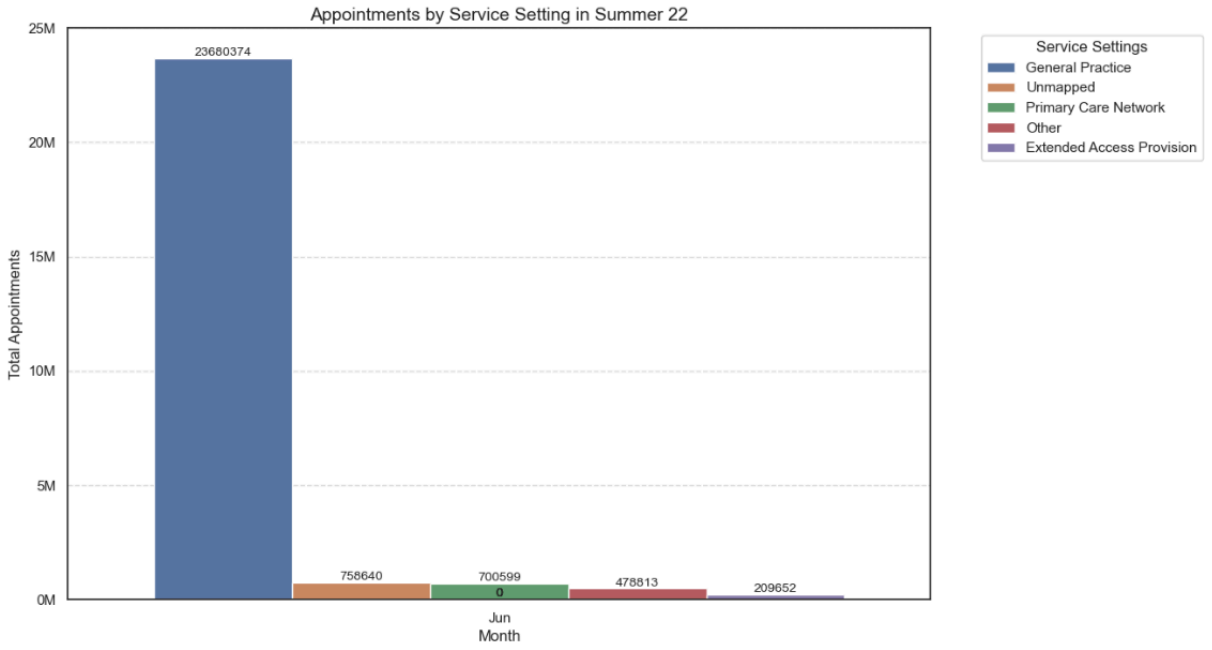
Appendix 12



Appendix 13



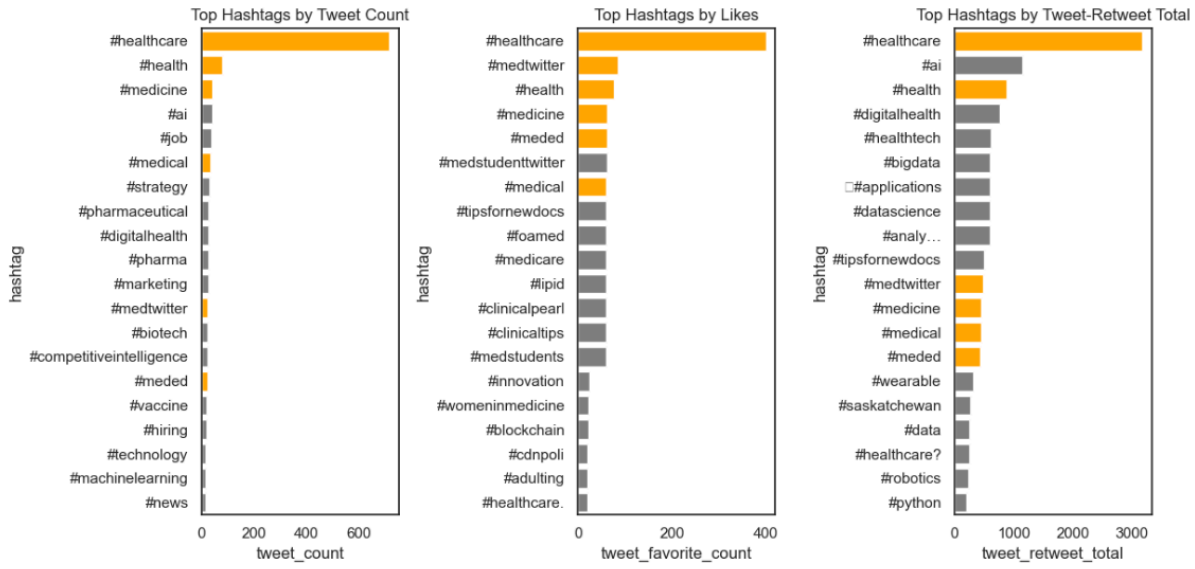
Appendix 14



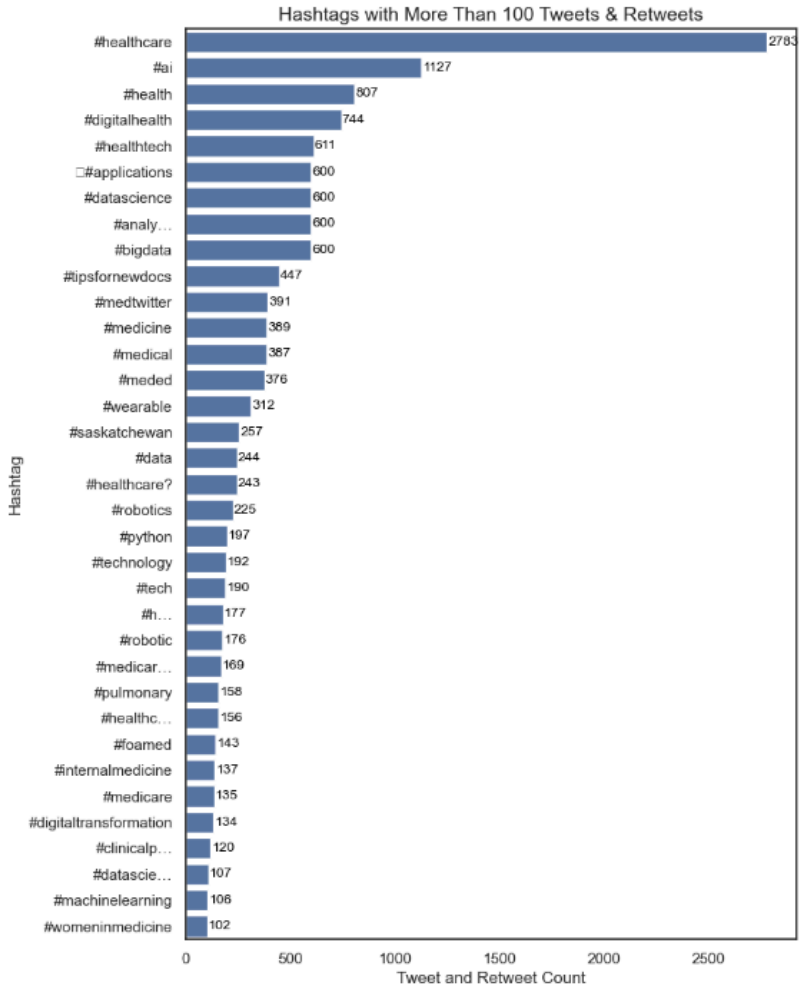
Appendix 15

	hashtag	tweet_count	tweet_retweet_count	tweet_favorite_count
0	#healthcare	716	2783.0	400.0
1	#health	80	807.0	76.0
2	#medicine	41	389.0	62.0
3	#ai	40	1127.0	13.0
4	#job	38	27.0	1.0
5	#medical	35	387.0	60.0
6	#strategy	30	29.0	0.0
7	#pharmaceutical	28	24.0	1.0
8	#digitalhealth	25	744.0	15.0
9	#pharma	25	25.0	0.0
10	#marketing	25	25.0	0.0
13	#competitiveintelligence	24	24.0	0.0
11	#medtwitter	24	391.0	85.0
12	#biotech	24	24.0	0.0
14	#meded	23	376.0	62.0
15	#vaccine	18	29.0	4.0

Appendix 16



Appendix 17



Appendix 18

```
[('#healthcare', 785), ('to', 581), ('the', 521), ('in', 353), ('and', 332), ('for', 286), ('a', 282), ('of', 268), ('rt', 192), ('is', 184), ('our',
163), ('-', 129), ('you', 129), ('with', 124), ('on', 123), ('we', 107), ('are', 103), ('at', 99), ('health', 97), ('this', 97), ('be', 93), ('your',
89), ('more', 89), ('&amp;', 86), ('healthcare', 85), ('how', 78), ('can', 74), ('care', 73), ('#health', 73), ('by', 72), ('as', 72), ('about', 71),
('from', 68), ('that', 68), ('learn', 57), ('it', 57), ('have', 54), ('their', 48), ('job', 47), ('will', 45), ('us', 45), ('what', 44), ('new', 43),
('join', 41), ('or', 40), ('help', 39), ('out', 39), ('an', 38), ('i', 37), ('click', 36)]
```

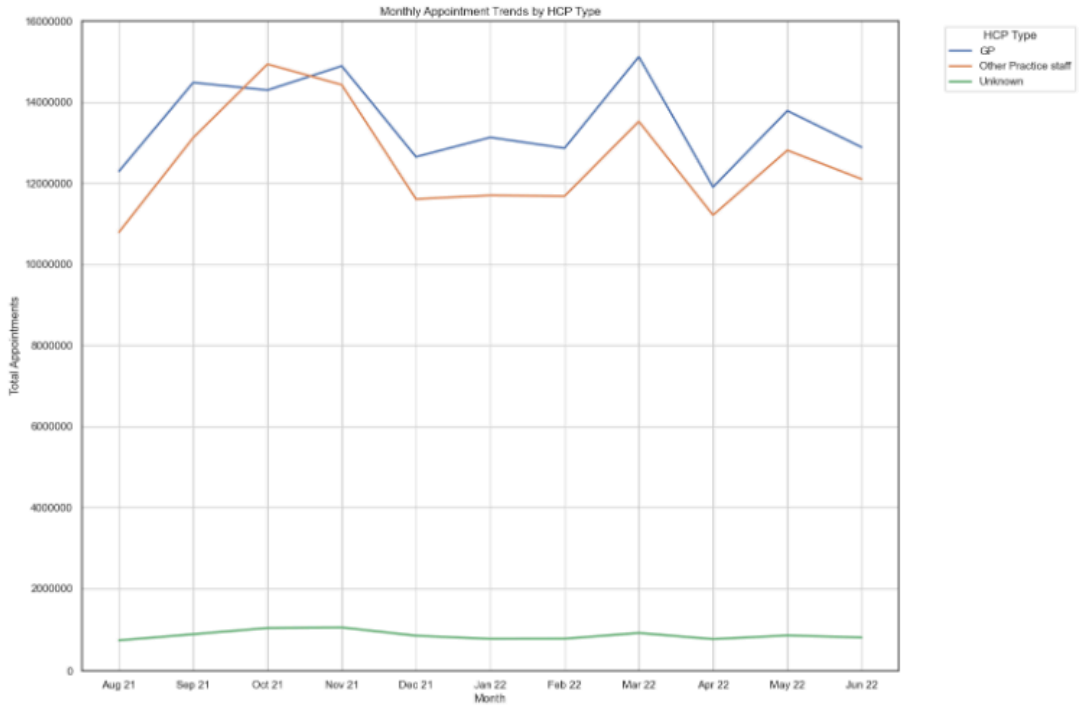
Appendix 19

	appointment_mode	count_of_appointments	percentage
0	Face-to-Face	183457603	62.0
2	Telephone	100237680	33.9
3	Unknown	8729565	2.9
1	Home Visit	1975204	0.7
4	Video/Online	1540555	0.5

Appendix 20

hcp_type appointment_mode	GP	Other Practice staff	Unknown	GP percentage
Face-to-Face	75149783	106401462	1906358	41.0
Home Visit	821878	890930	262396	41.6
Telephone	70152247	28334498	1750935	70.0
Unknown	1476503	1497787	5755275	16.9
Video/Online	727233	809890	3432	47.2

Appendix 21



Appendix 22

	time_between_book_and_appointment	count_of_appointments	percentage
6	Same Day	130964160	44.3
2	2 to 7 Days	60720229	20.5
4	8 to 14 Days	37490409	12.7
0	1 Day	25854059	8.7
1	15 to 21 Days	19437514	6.6
3	22 to 28 Days	11417519	3.9
5	More than 28 Days	9868508	3.3
7	Unknown / Data Quality	188209	0.1

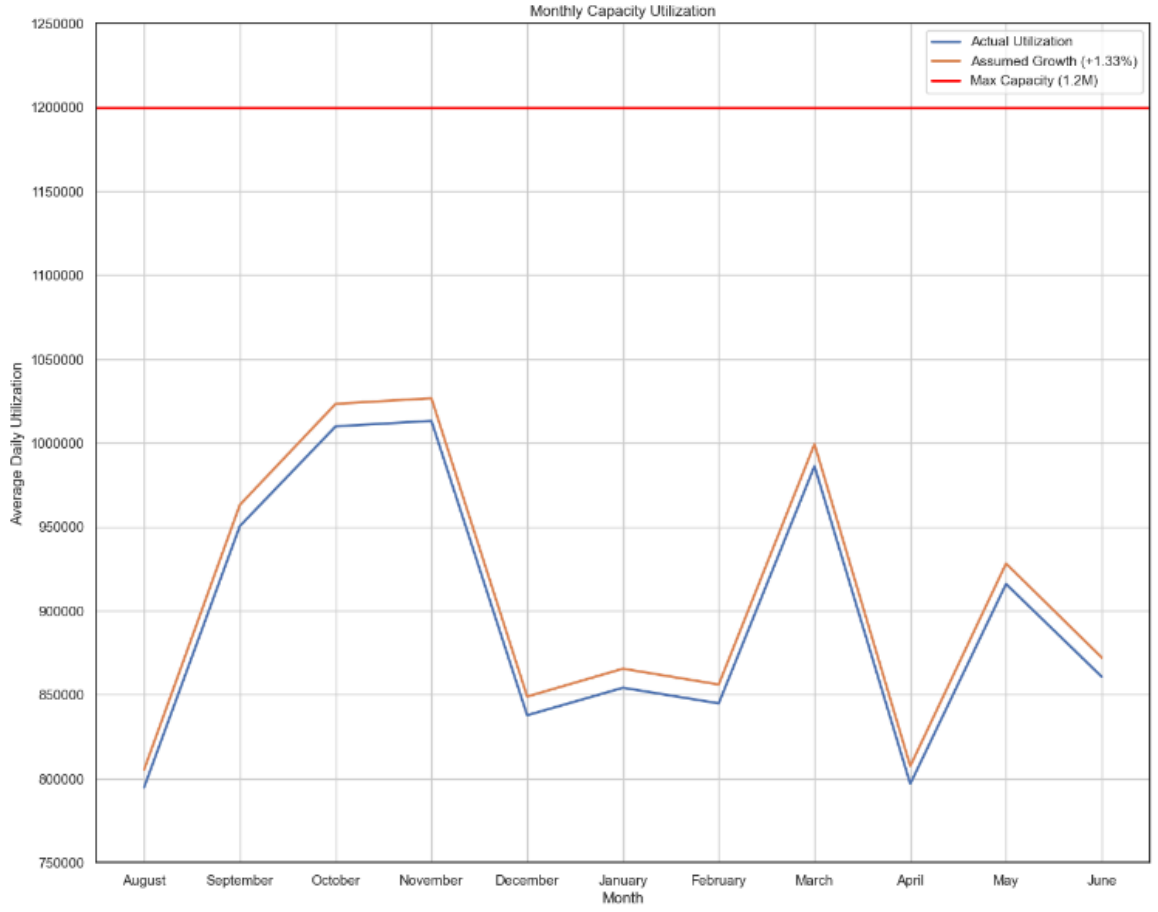
Appendix 23

time_between_book_and_appointment	Same Day	1 Day	15 to 21 Days	2 to 7 Days	22 to 28 Days	8 to 14 Days	More than 28 Days	Unknown / Data Quality
appointment_mode								
Face-to-Face	35.9	8.8	7.9	23.1	4.7	15.2	4.4	0.1
Home Visit	56.5	14.4	2.2	17.6	1.4	5.5	1.6	0.7
Telephone	58.9	8.5	4.3	16.0	2.4	8.4	1.4	0.1
Unknown	51.1	8.4	5.2	18.8	3.0	10.3	3.2	0.1
Video/Online	33.5	14.8	7.3	22.5	4.3	13.3	4.3	0.1

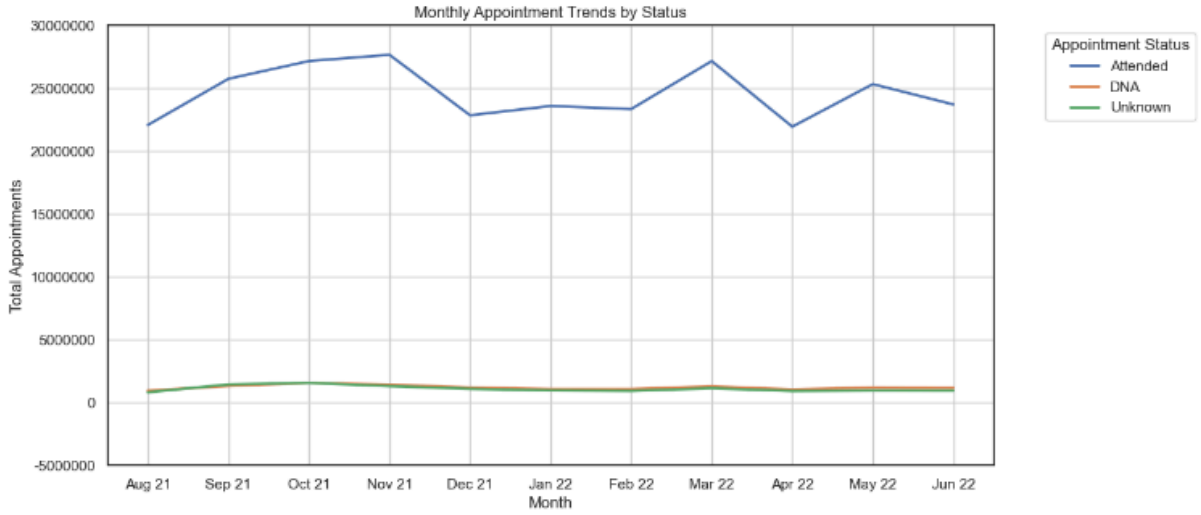
Appendix 24

	appointment_year	appointment_month	count_of_appointments	avg_daily_utilization
0	2021	August	23843177	794772.6
1	2021	September	28514685	950489.5
2	2021	October	30296850	1009895.0
3	2021	November	30395923	1013197.4
4	2021	December	25132174	837739.1
5	2022	January	25623928	854130.9
6	2022	February	25344812	844827.1
7	2022	March	29586020	986200.7
8	2022	April	23904960	796832.0
9	2022	May	27478652	915955.1
10	2022	June	25819426	860647.5

Appendix 25



Appendix 26



Appendix 27

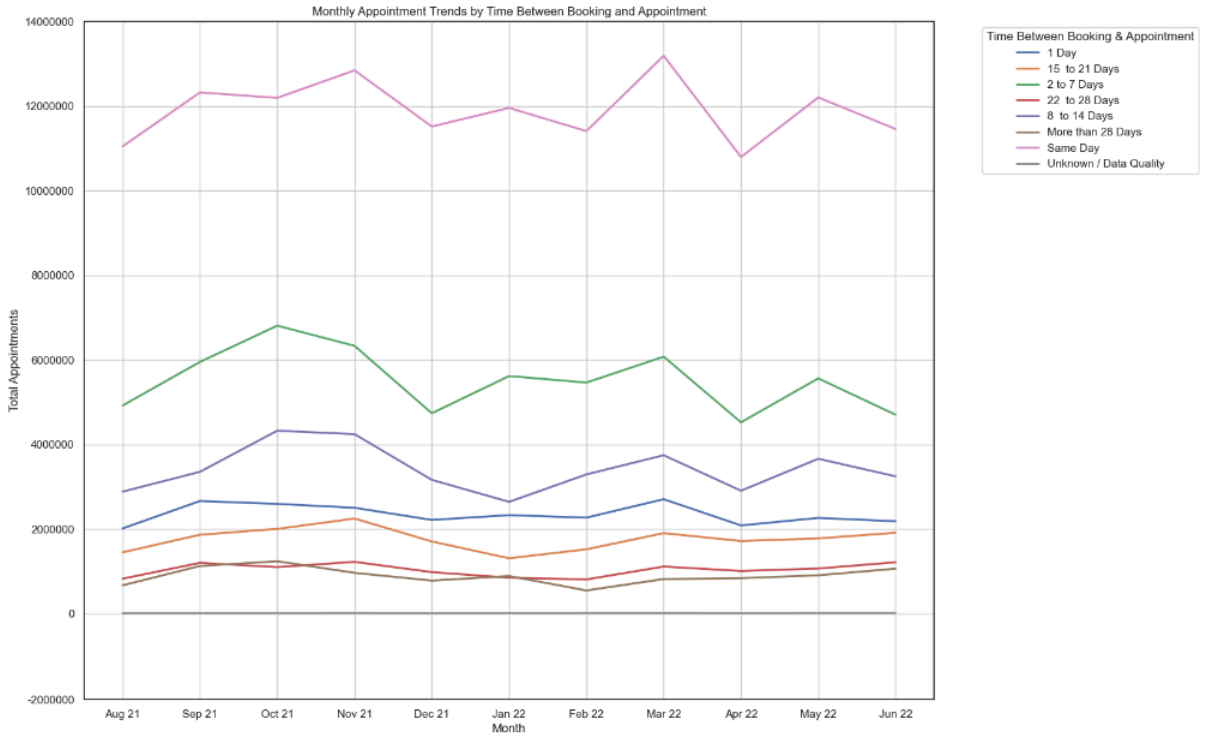
appointment_status	Attended	DNA	Unknown	Total	Attended (%)
time_between_book_and_appointment					
Same Day	125796242	2284169	2883749	130964160	96.1
1 Day	24003592	1039523	810944	25854059	92.8
2 to 7 Days	54636124	3689732	2394373	60720229	90.0
8 to 14 Days	32523007	2864860	2102542	37490409	86.8
15 to 21 Days	16506091	1563712	1367711	19437514	84.9
22 to 28 Days	9474372	931458	1011689	11417519	83.0
More than 28 Days	7539601	907394	1421513	9868508	76.4
Unknown / Data Quality	138394	4948	44867	188209	73.5

Appendix 28

appointment_status	Attended	DNA	Unknown	Total	Attended (%)
appointment_mode					
Telephone	95350959	2081545	2805176	100237680	95.1
Video/Online	1414513	76820	49222	1540555	91.8
Unknown	7838384	304109	587072	8729565	89.8
Face-to-Face	164507523	10747016	8203064	183457603	89.7
Home Visit	1506044	76306	392854	1975204	76.2

Appendix 29

Appendix 30



Appendix 31

time_between_book_and_appointment icb_ons_code	1 Day	15 to 21 Days	2 to 7 Days	22 to 28 Days	8 to 14 Days	More than 28 Days	Same Day	Unknown / Data Quality	Total Appointments
E54000050	9.1	5.8	21.8	3.4	12.5	3.0	44.4	0.0	43054121
E54000054	9.6	5.9	19.3	3.5	11.2	4.0	46.5	0.0	36075020
E54000057	9.5	6.1	20.5	3.6	11.7	2.8	45.7	0.0	34003953
E54000008	10.2	5.1	20.1	2.9	10.8	2.2	48.7	0.0	33064205
E54000027	11.5	4.6	24.0	2.2	11.5	1.5	44.7	0.1	29380767
E54000029	12.0	3.9	23.6	2.0	9.9	1.3	47.1	0.3	23594259
E54000042	8.8	6.4	20.2	3.7	12.4	3.3	45.1	0.1	22968118
E54000051	9.0	6.0	22.1	3.7	12.6	3.5	42.9	0.0	22964301
E54000032	9.1	6.3	20.4	3.9	12.1	3.5	44.7	0.0	22623978
E54000053	8.3	6.4	19.9	4.1	12.2	4.0	45.1	0.0	22566486
E54000044	8.4	5.8	20.7	3.4	11.8	2.6	47.3	0.0	22546031
E54000048	8.7	6.2	20.2	3.8	11.8	2.8	46.4	0.1	21487277
E54000030	11.3	4.6	22.1	2.4	10.6	1.5	47.4	0.2	20241452
E54000061	8.7	6.5	19.9	4.1	12.6	4.8	43.3	0.0	19282941
E54000037	9.1	5.9	21.1	3.9	12.2	3.9	43.8	0.0	19139882
E54000031	9.7	4.9	22.6	2.5	11.6	1.5	47.2	0.0	17972966
E54000025	8.5	5.9	19.7	3.3	11.7	2.5	48.4	0.0	17684743
E54000062	10.3	4.4	21.0	2.4	10.1	2.4	49.3	0.1	17514050