THE NHS: 2026 & BEYOND

Exploring NHS data and social media for data driven insights

The UK population is growing. NHS infrastructure will need investment to grow relative to demand and the organisation will need to make better use of its existing resources. This project explores available data to inform strategic decisions that shape capacity planning and optimise service delivery.

Toby Draper

Table of Contents

1. Introduction

Due to concerns within NHS leadership about the financial impact of missed appointments and the service's readiness to meet increasing demand, this project was initiated to achieve the following objectives:

- The primary objectives are to answer:
  - Has there been adequate staff and capacity in the networks?
  - What was the actual utilisation of resources?
- Refine other stakeholder questions into actionable analytical tasks
- Provide recommendations to inform NHS strategy

This document is a guide for technical stakeholders, accompanying:

- A video presentation summarising key findings - **Draper_Toby_DA201_Assignment_Presentation.mp4**
- **Draper_Toby_DA201_Assignment_Notebook.ipynb** containing code used to analyse data
  - Section numbering in this document directly correspond to the notebook structure
  - Markdown cells and inline comments guide technical stakeholders through code execution
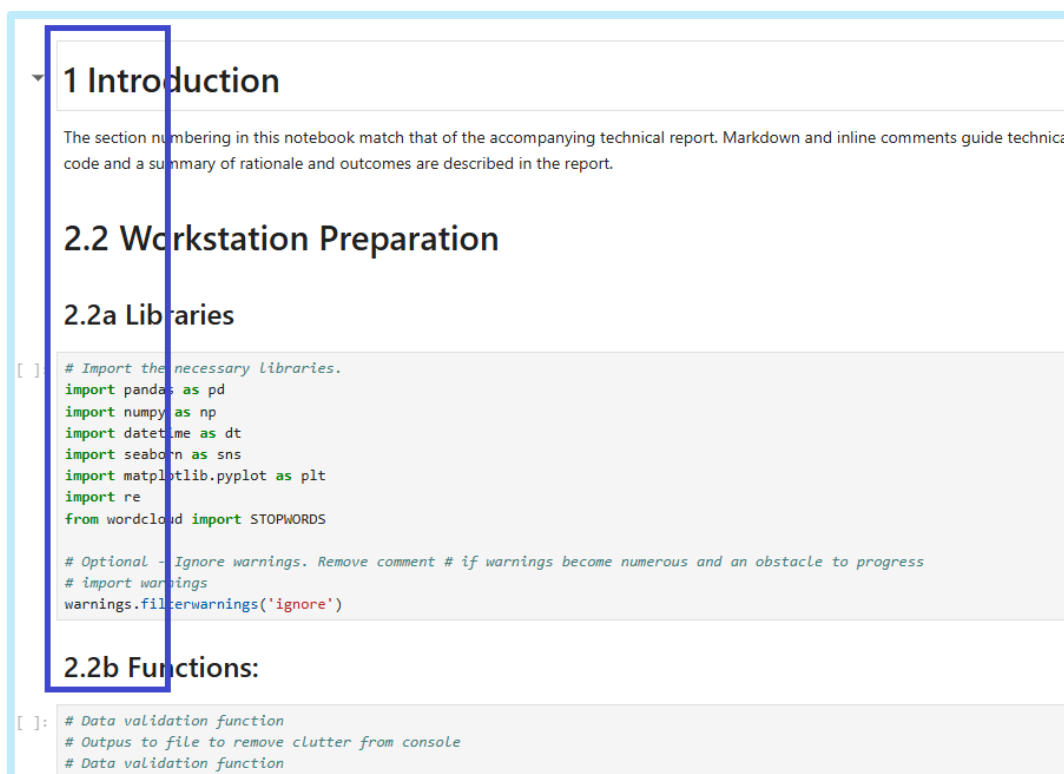

Figure 1: Section Numbering in Jupyter Notebook

2. Analytical Approach

2.1 Project materials

NHS leadership provided;

Filename	File type	Description	Comments
actual_duration.csv	Comma separated values	Counts of appointments for each appointment duration category for all NHS locations	Useful for obtaining full breakdown of NHS geography at three levels of hierarchy (Region, ICB, SICBL)
appointments_regional.csv	Comma separated values	Counts of General Practice appointments by Region	Useful for identifying whether appointments are attended, whether the appointment was in person or a virtual meeting and identifying time elapsed from booking
national_categories.xlsx (see comment and Appendix A)	**MS Excel**	**Counts of appointments defined in framework of three levels of categorisation**	**The file used in the accompanying notebook and shared in the Google Drive linked below has been converted to .csv format for faster ingestion**
metadata_nhs.txt	Text file	Descriptions for the three files containing appointment data	Contains exhaustive information, likely for auditing/compliance and of limited use for data analysts. Refer to validation routines to obtain metadata
tweets.csv	Comma separated values	A selection of tweets that potentially indicate relevant social sentiment about aspects of the service	Assess data to inform business case for or against paid access to tweet data

Backup of project files: https://drive.google.com/drive/folders/1Hjp8fjY3wsIhdasSu9KL-pyyxZfm-lTy?usp=sharing

2.2 Workstation Preparation

a. Libraries imported to the workflow:

Library	Used for	Project data types	Comment
pandas	Data wrangling	Appointment data	Standard data analysis library
numpy	Numeric operations	Appointment data	Standard data analysis library
datetime	Date manipulation	Appointment data	Parse, format and compare date fields
seaborn	Visualisations	Appointment data	Good for statistical charts
matplotlib	Visualisations	Appointment data	Chart formatting used in tandem with seaborn
re	Text cleaning	Tweet data	Regex-based filtering and pattern matching in text
WordCloud	WordCloud generation	Tweet data	Convert text into visual frequency map
STOPWORDS	Stopword filtering	Tweet data	Remove common "filler words"

b. Functions called in the workflow:

Function	Used for	Details	Comment
log_data_validation	Data validation	Appointment data	File: validation_log_1
log_tweet_validation	Data validation	Tweet data	File: validation_log_2
get_season	Assigns a season label based on numeric month	Conditional logic (if, elif, else) to populate new season column in ar and nc,	Called once when wrangling data for seasonal analysis
export_countplot_short	Small countplots	Visualisations used to present verification of record counts in section 2.4, reporting Service Settings, Context Types and Appointment Statuses	For countplots when column count <= 10
export_countplot_long	Larger countplots	Used to visualise record count of National Categories in the nc data set	One countplots had column count >= 10 in section 2.4
export_piechart_with_legend	Breakdown appointment data by % in categories (e.g. Service Setting)	For pie charts with percentage labels and external legend	Used to summarise Service Settings at national and local level which appear in the presentation
export_lineplot_with_legend	Monthly trend analysis	Time series with month on x axis, record count on y axis and categories in hue settings (Service Setting, Context Type etc)	Service Utilisation & Attendance Analysis slides in presentation
export_barplot_with_legend	Seasonal analysis	It was difficult to see seasonal trends in lineplots. Opted to use barplots with hue set to season to emphasise the season a month falls in	Used to produce barplots that appear in the presentation

2.3 Data Ingestion & Validation Checks

Full validation summaries can be reviewed in corresponding Validation Logs. Also, see Appendix B: Example of Validation Log Output

DataFrame	Row Count	Column Count	Comments	Validation Log
ad	137793	8	No validation issues identified	validation_log_1.txt
ar	596821	7	No validation issues identified	validation_log_1.txt
nc	817394	8	No validation issues identified	validation_log_1.txt
tweets	1174	10	No validation issues identified	validation_log_2.txt

- The data sets appear to be pre-cleaned. No additional preparation required
- Data sets not joined. Row counts mismatch significantly, also uncertainty regarding join key
- tweets was validated separately due to difference in data type and reduced record count

2.4 Data Wrangling & Exploratory Analysis

Further validation/exploration of the data sets assessed whether they are adequate to produce reliable outputs. Reflected in stakeholder questions;

- *What is the number of locations (a), service settings, context types, national categories, and appointment statuses (b) in the data sets?*
- *What is the date range of the provided data sets (c), and which service settings reported the most appointments for a specific period (d)?*
- *What is the number of appointments and records per month (e)?*

a. Regional Summary

Verify NHS England geography (Figure 2) against the ad data set (Figure 3).

English health geography

Seven NHS England (Regions) (NHSER) cover all of England, within which there are 42 Integrated Care Boards (ICB). At the primary care level, there are 106 Sub Integrated Care Board Locations (SICBL).

Figure 2: NHS Regional Structure. Source: Office for National Statistics [ONS], n.d.

region_ons_code	icb_ons_code	sub_icb_location_ons_code
E40000003	5	5
E40000005	6	11
E40000006	7	7
E40000007	6	14
E40000010	3	27
E40000011	11	19
E40000012	4	23
Total	42	106

Figure 3: NHS Regional Structure

See Appendix C: NHS Geography Detail from Appointment Data for full breakdown of NHS geography

b. Operational Summary

Create countplots "Service Settings", "Context Types", "National Categories" and "Appointment Statuses" (Figures 4– 7).
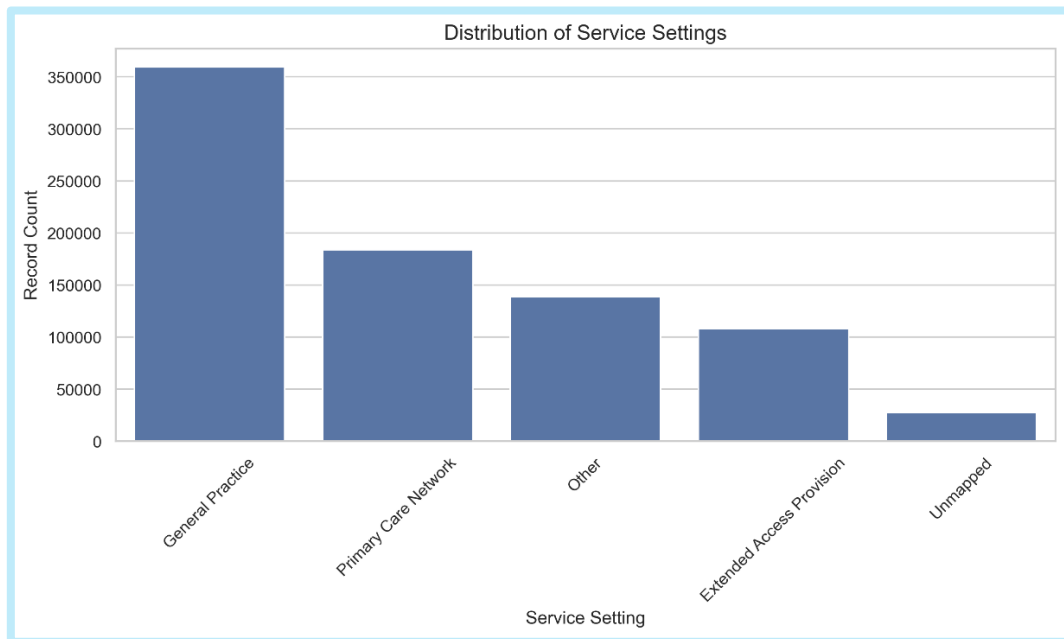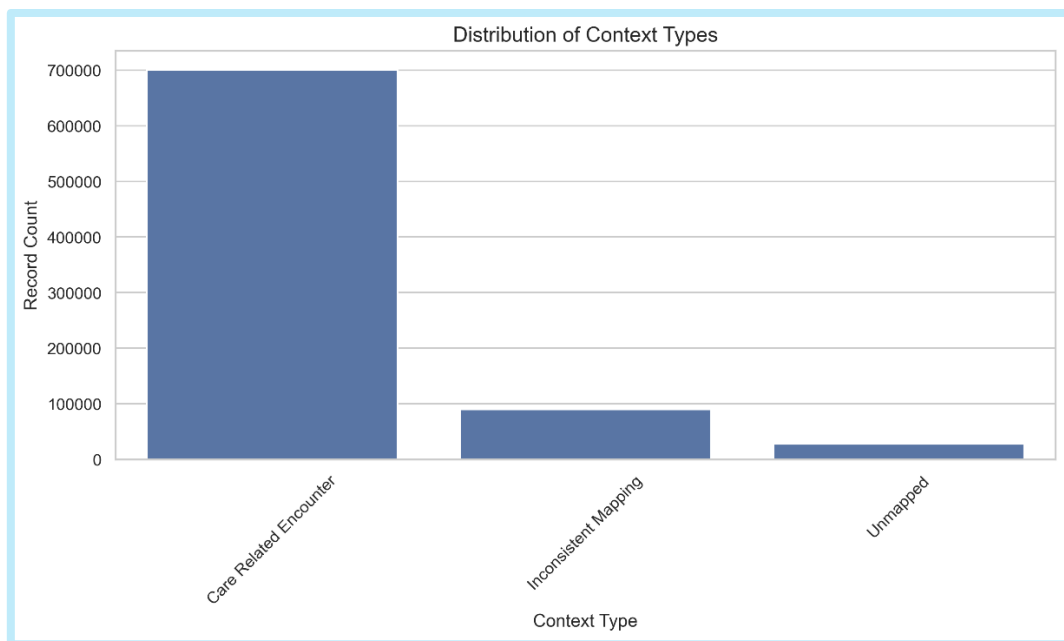

Figure 4: Service Settings (nc)
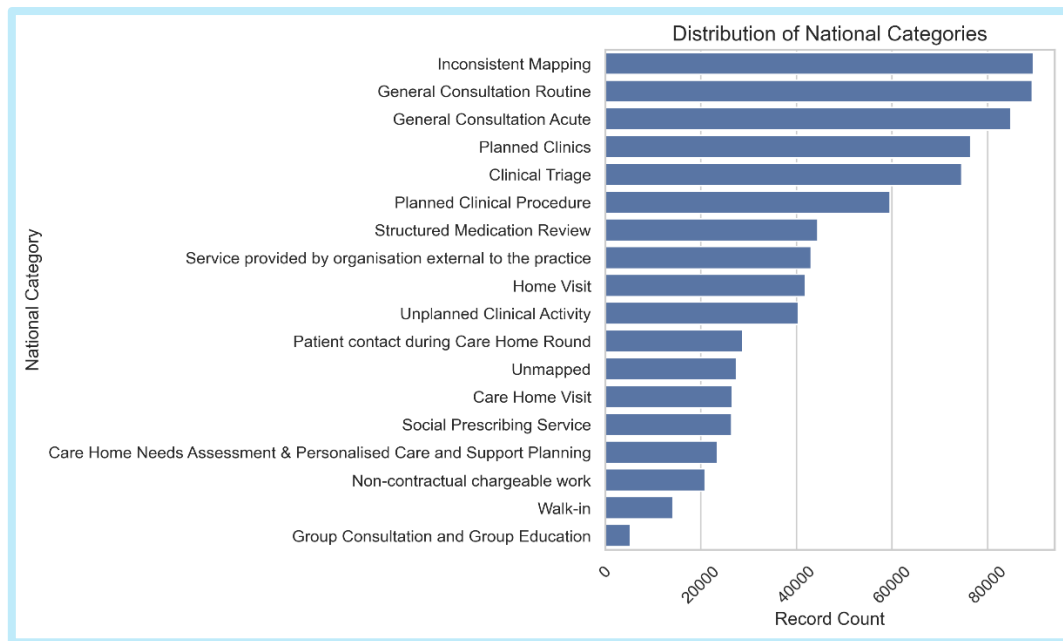

Figure 5: Context Types (nc)

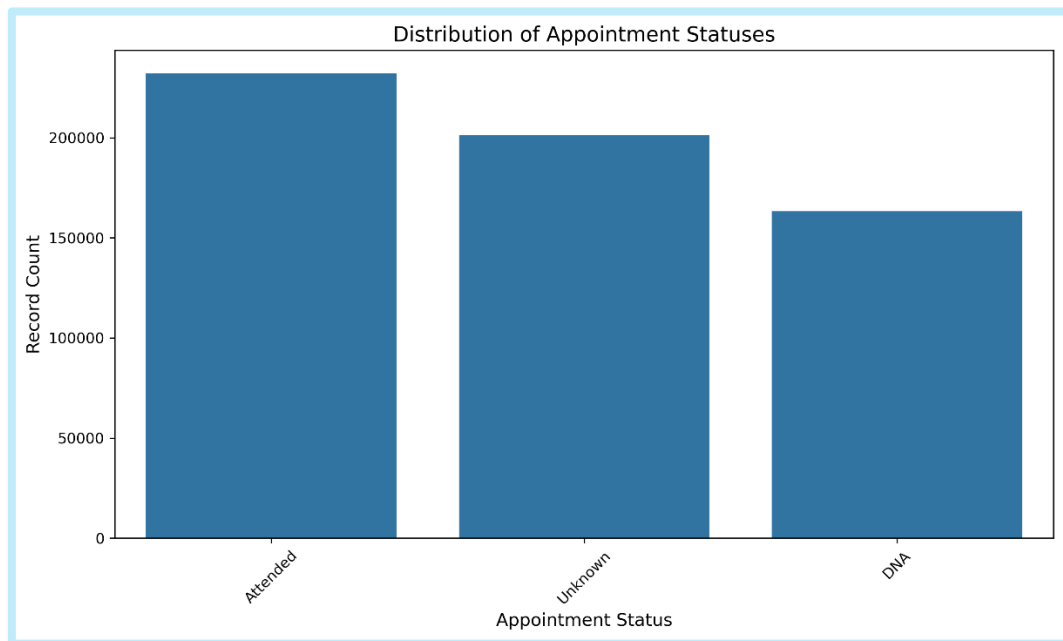**Figure 6: National Categories (nc)**

Figure 7: Appointment Statuses (ar)

c. Date Summary

Ensure dates comply with UK Government Standards (Figure 8).
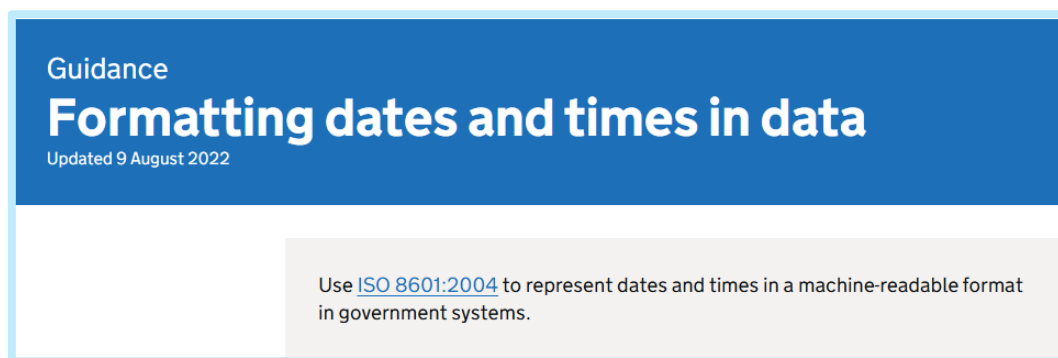

Figure 8: Date and Time Formatting Standard. Source: Cabinet Office [GOV.UK], 2022

Date format conversion;

- ad and nc included full dates in ISO 8601 YYYY-MM-DD format. These were converted to datetime using format inference. Checks for invalid entries found none.
- ar contained dates in YYYY-MM format. Converted to Period dtype to retain structure and enable analysis.

Summary of changes and date ranges:

df	Dtype	Column	Min	Max	Count
ad	datetime64	appointment_date	2021-12-01	2022-06-30	8
ar	Period	appointment_month	2020-01	2022-06	30
nc	datetime64	appointment_date	2021-08-01	2022-06-30	11

d. Service Setting Summary

The underlying totals for this chart appear in Appendix D: Appointment Totals by Service Setting.
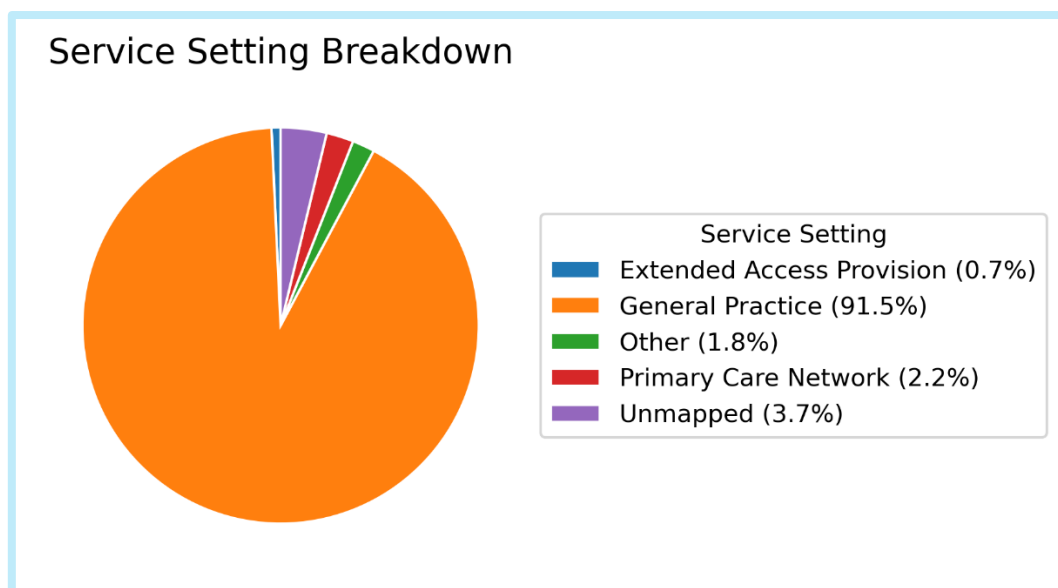

Figure 9: Service Settings by patient demand (nc)

e. Appointments Monthly Summary

	df	Monthly Average	Top Month	Top Month Total
Records	ar	19894	2020-01	20889
Appointments	ar	24760150	2021-11	30405070

More detail from the subsets used to generate these figures appears in Appendix E: Appointment Statistics.

2.5 Data Wrangling Steps for Visualisations

a. Seasonal Analysis

Steps to prepare the visualisations for monthly/seasonal analysis:

df name	Source df	Description	Comment
nc_date	nc	Created for visualising monthly analysis/trends in nc	Convert the dtype of the column appointment_month to string to allow for easier plotting
nc_date_ss	nc_date	Created for visualising monthly analysis/trends of Service Settings	Used in time series lineplots
nc_date_ct	nc_date	Created for visualising monthly analysis/trends of Context Types	Used in time series lineplots
nc_date_nc	nc_date	Created for visualising monthly analysis/trends of National Categories	Used in time series lineplots
nc_season	nc_date	Created to add a 'season' column to nc_date	'season' column populated by function get_season(month)
nc_season_counts	nc_season	Aggregate appointment counts by month and season to prepare seasonal trend data	Used in a barplot displaying seasonal trends
nc_season_totals	nc_season_counts	Aggregate to create high level summary of seasonal totals	See Appendix F & G: Monthly/Seasonal Totals in nc
ar_season	ar	Created to add a 'season' column to ar	'season' column populated by function get_season(month)
ar_season_counts	ar_season	Aggregate appointment counts by month and season to prepare seasonal trend data	Used in a barplot displaying seasonal trends over a wider date range
ar_season_totals	ar_season_counts	Aggregate to create high level summary of seasonal totals	See Appendix H: Seasonal Totals in ar

b. Social Media Analysis

Steps to prepare the visualisations for social media analysis:

df name	Source df	Description	Comment
tweets_text	tweets	Create a new DataFrame containing only the text	Extract text to build Wordcloud – step #1
text	tweets_text	Convert all text to strings and merge into one block	Wordcloud wrangling step #2
text	tweets_text	Remove punctuation and numbers	Wordcloud wrangling step #3
text	tweets_text	Convert text to lower case	Wordcloud wrangling step #4
text	tweets_text	Remove stopwords from text (the, and, is etc)	Wordcloud wrangling step #5. Data ready for visualisation

c. Local Highlights

Steps to prepare the visualisations for analysis of SIBL London NW:

df name	Source df	Description	Comment
nc_local	nc	Isolated data for SIBL London NW	
nc_local_count	nc_local	Group and aggregate appointment counts by service setting	Repeat process used to create pie chart in 2.4e Appointments Monthly Summary
nc_local_month	nc_local	Aggregate on monthly level and determine the sum of records per month for service_setting	Repeat process used to create time series chart in 2.5a Seasonal Analysis

d. Service Utilisation & Appointment Analysis

Steps to prepare the visualisations for analysis of service utilisation and appointment categorisation:

df name	Source df	Description	Comment
ar_agg	ar	Create a new DataFrame to prepare for categorical analysis of ar	Used in 5 time series visualisations to show monthly trends for the various ways appointments are categorised
ar_agg	ar	Change the data type of the appointment month to string	Allow for easier plotting
ar_df	ar	Create a new DataFrame where utilisation will be calculated	Used in 2 series visualisations to show monthly utilisation trends
ar_df	ar	Calculate utilisation	Divide the sum of the monthly appointments by 30 to get a daily value
ar_df	ar	Populate a new column with a benchmark value as basis for monthly utilisation calculations	NHS provided figure of 1.2 million appointments per day for planning purposes
ar_df	ar	Calculate the gap between actual utilisation and the NHS benchmark	
ar_df	ar	Reshape data, stacking selected columns into 'metric' for use as hue in lineplot	Use melt so column headers are now entries in the 'metric' column. Used in 1 visualisation.
status_totals	ar_agg	Generate readable totals, shared in Appendix J	Figures referenced in presentation and used for annotation
ar_unattended	ar_agg	Create DataFrame to analyse unattended appointments to map the time between booking and appointment date	Deeper analysis of unattended appointments to identify whether delays between booking and appointment are a factor in non-attendance

3. Visualisations

The visualisations used in this project fall broadly into the following categories;

- Surface level insights described in section 2.3, created to deliver statistical information efficiently in this report
- Seasonal analysis to identify emerging trends
- Local highlights – a study of SICBL London NW
- Social media analysis, evaluating extracted Tweet data
- Service utilisation and appointment analysis

3.1 Seasonal Analysis

This analysis is originated with the following stakeholder question:

- *What monthly and seasonal trends are evident, based on the number of appointments for service settings, context types, and national categories?*

Summary of visualisations created for Seasonal Analysis

df	Function	Title	Description	Date Range	Presented?
nc_date	export_lineplot_with_legend	Monthly Appointments	Time series to gain monthly trends for all appointments	August 2021 to June 2022	Y
nc_date_ss	export_lineplot_with_legend	Monthly Appointments - Service Settings	Time series to gain monthly trends for Service Settings	August 2021 to June 2022	Y
nc_date_ct	export_lineplot_with_legend	Monthly Appointments - Context Type	Time series to gain monthly trends for Context Types	August 2021 to June 2022	N
nc_date_nc	export_lineplot_with_legend	Monthly Appointments - National Category	Time series to gain monthly trends for National Categories	August 2021 to June 2022	N
nc_season_counts	export_barplot_with_legend	Seasonal Trend - All Categories	**A barplot using hue to highlight seasons results in coloured bars which are much more effective for seasonal comparison than a line plot**	August 2021 to June 2022	Y
ar_season_counts	export_barplot_with_legend	Seasonal Trend - Extended Range	Drill down into GP activity (95% of appointments). Covers wider date range to show genuine seasonal trends	January 2020 to June 20222	Y

3.2 Social Media Analysis

Study instigated by stakeholder question:

- *What are the top trending hashtags (#) on the supplied Twitter data set and how can this be used in the decision-making process?*

Limiting factors were identified after analysing hashtag data, which were found to contain very generic terms (Appendix I: Social Media Analysis).

A WordCloud was chosen to visualise all tweet text, a design choice intended to add emphasis to the recommendations drawn from this section of analysis.

Figure 10: Wordcloud visualisation containing Tweet data

3.3 Local Highlights

Summary of visualisations created for focus on SICBL London NW

df	Function	Title	Description
nc_local_count	export_piechart_with_legend	Service Setting Breakdown - NW London	Pie chart for service setting breakdown in NW London
nc_local_month	export_lineplot_with_legend	Monthly Appointments - NW London	Report Monthly trends by service setting just for this location

3.4 Service Utilisation & Appointment Analysis

Associated stakeholder questions:

- *What was the actual utilisation of resources?*
- *What insights can be gained by looking at missed appointments?*
- *What are the most important patterns visible in the data relating to the use case?*

Summary of visualisations created:

df	Function	Title	Description	Date Range	Presented?
ar_df	export_lineplot _with_legend	Appointments by Month	Lineplot indicating number of monthly visits	January 2020 – June 2022	Y
plot_df	export_lineplot _with_legend	Utilisation Analysis	Compare monthly utilisation against NHS benchmark – 1.2 million daily appointments	January 2020 – June 2022	Y
ar_agg	export_lineplot _with_legend	HCP Types by Month	Monthly comparison of appointments categorised by Health Care Professional type	January 2020 – June 2022	N
ar_agg	export_lineplot _with_legend	Attendance Status by Month	Monthly attendance trends	January 2020 – June 2022	Y
ar_agg	export_lineplot _with_legend	Appointment Mode by Month	Monthly comparison of appointments categorised by Appointment Mode	January 2020 – June 2022	Y
ar_agg	export_lineplot _with_legend	Booking Interval	Evaluate evidence of trends in duration between when the appointment was booked and the date of the appointment	January 2020 – June 2022	Y
ar_unattended	export_lineplot _with_legend	Booking Interval of Unattended Appointments	Deep dive into unattended appointments to investigate relationship between unattendance and the length of time elapsed since the booking	January 2020 – June 2022	Y

4. Trends, Insights & Recommendations

Summary of insights gathered, trends observed and all recommendations made to influence NHS strategy for capacity and service delivery planning, linked to stakeholder question:

- *What insights can be gained from the data, and what recommendations can be made to the NHS based on these insights?*

4.1 Developing Actionable Analytical Tasks

These tasks were developed in response to the stakeholder questions responded to throughout this document. They should be repeated and developed further in future cycles.

Stakeholder Question	Analytical Task	Deliverable
What is the number of locations, service settings, context types, national categories, and appointment statuses in the data sets?	Interrogate data sets ad, ar, nc ensure quality, record count, categorisation correct Validate against other sources where available	Visualisations shared in section 2.4 in Figures 4-7
What is the date range of the provided data sets, and which service settings reported the most appointments for a specific period?	Ensure dates comply with UK Government Standards, establish the date ranges. Report on Service Settings to validate accuracy.	Refer to section 2.4c for details of date ranges and formats. 2.4d for service setting report.
What is the number of appointments and records per month?	Establish volume metrics to check accuracy of visualisations produced by the project	2.4e record and appointment counts
What monthly and seasonal trends are evident, based on the number of appointments for service settings, context types, and national categories?	Visualise monthly and seasonal trends to identify periods of high demand	See list of visualisations produced in section 3.1
What are the top trending hashtags (#) on the supplied Twitter data set and how can this be used in the decision-making process?	Analyse social media data for trends that could inform project insights	Section 3.2 contains an example visualisation. Alternatives in Appendix H
Was there adequate staff and capacity in the networks?	Measure utilisation against benchmarks provided by NHS	Time series chart, title Utilisation Analysis, section 3.4
What was the actual utilisation of resources?	Measure appointment counts by month (ar) for all appointments and broken down by appointment categories	Time series charts of HCP Type, Appointment Status, Appointment Mode, Time Between Book and Appointment in section 3.4
What insights can be gained by looking at missed appointments?	Analyse missed appointments in ar to analyse impact and identify improvement plans	Refer to report "Booking Interval of Unattended Appointments", section 3.4
What are the most important patterns visible in the data relating to the use case?	Evaluate visualisations produced for discernible patterns	Summary in section 4.2 and Slide x of presentation
What insights can be gained from the data, and what recommendations can be made to the NHS based on these insights?	Interpret the patterns observed in previous step into concise insights and recommendations and deliver to stakeholders	Summary in sections 4.2 and Slide x of presentation

4.2 Trends & Insights

Summary of patterns and observations shared in presentation

- General Practice 95% of all appointments
- Appointments peaked in October 2022 (x million appointments - nc)
- Higher demand in Autumn
- Extended range of dates (GP Appointments – ar) confirm Autumn peak
- Tweet data provided provides no insights about missed appointments
- Utilisation trending towards NHS benchmark (1.2 million appointments per day)
- 91% of appointments are attended by patients
- Telephone appointments became popular during first Covid lockdown
- As time advances from time of booking, likelihood of unattendance rises

4.3 Recommendations

a. Main objectives

1. Has there been adequate staff and capacity in the networks?
    - Yes – metrics show sufficient daily capacity
    - Actual utilisation trending towards planning benchmark 1.2m appointments per day
    - **Action: Monitor trend and calculate new daily appointment benchmark accordingly**
    - **Action: Plan staffing levels to new benchmark**
2. What was the actual utilisation of resources?
    - Peaked in 2022-10/11
    - **Action: Ensure demand can be met during Autumn peak**

b. Additional recommendations

- **Action: Obtain data for period between 2022-06 and present for review**
- Optimise service capacity and reduce costs by addressing unattended appointments
    - **Action: Explore increasing proportion of telephone appointments to reduce "unattended same day" appointments**
    - **Action: Reduce time between booking and appointment for face to face appointments**
    - Social media campaigns could be leveraged to promote guidelines to patient booking
    - Tracking such a campaign would potentially make investment in access to social media data worthwhile

5. References

1. English Health Geography (Office for National Statistics)

https://www.ons.gov.uk/methodology/geography/ukgeographies/healthgeography#:~:text=Seven%20NHS%20England%20(Regions)%20(,UK%20Geography%20for%20further%20details.

2. Guidance: Formatting dates and times in data (Gov.UK - Open standards for government)

https://www.gov.uk/government/publications/open-standards-for-government/date-times-and-time-stamps-standard

6. Appendix

Appendix A: File format of national_categories.xlsx data set

Preparing for this project required repeated loads of the provided files into DataFrames in Jupyter Notebooks. It was observed that ingestion of national_categories always took much longer than other files. Assuming this was due to its MS Excel file format, the file was converted to comma separated values (.csv) as a test. The resulting file loaded much more quickly as .csv and so a converted national_categories.csv has been shared in the Google drive linked to in the report.

```python
import pandas as pd

nc = pd.read_excel('national_categories.xlsx')

nc.to_csv('national_categories.csv')
```

Appendix B: Example of Validation Log Output

=== Validation Log For Actual Duration ===

=======================================

Shape:

(137793, 8)

Non-Null Counts:

sub_icb_location_code   137793

sub_icb_location_ons_code   137793

sub_icb_location_name   137793

icb_ons_code   137793

region_ons_code   137793

appointment_date   137793

actual_duration   137793

count_of_appointments   137793

dtype: int64

Empty Columns:

Index([], dtype='object')

Data Types:

sub_icb_location_code   object

sub_icb_location_ons_code   object

sub_icb_location_name   object

icb_ons_code   object

region_ons_code   object

appointment_date   object

actual_duration   object

count_of_appointments   int64

dtype: object

Unique Values:

sub_icb_location_code   106

sub_icb_location_ons_code   106

sub_icb_location_name   106

icb_ons_code   42

region_ons_code   7

appointment_date   212

actual_duration   7

count_of_appointments   7611

dtype: int64

Duplicate Values:

0   False

1   False

2   False

3   False

4   False

        ...

137788 False

137789 False

137790 False

137791 False

137792 False

Length: 137793, dtype: bool

Descriptive Statistics:

        count_of_appointments

count   137793.000000

mean   1219.080011

std   1546.902956

min   1.000000

25%   194.000000

50%   696.000000

75%   1621.000000

max   15400.000000

Head:

 sub_icb_location_code sub_icb_location_ons_code \

0	00L	E38000130
1	00L	E38000130
2	00L	E38000130
3	00L	E38000130
4	00L	E38000130

               sub_icb_location_name icb_ons_code region_ons_code \

0	NHS North East and North Cumbria ICB - 00L	E54000050	E40000012
1	NHS North East and North Cumbria ICB - 00L	E54000050	E40000012
2	NHS North East and North Cumbria ICB - 00L	E54000050	E40000012
3	NHS North East and North Cumbria ICB - 00L	E54000050	E40000012
4	NHS North East and North Cumbria ICB - 00L	E54000050	E40000012

	appointment_date	actual_duration	count_of_appointments
0	01-Dec-21	31-60 Minutes	364
1	01-Dec-21	21-30 Minutes	619
2	01-Dec-21	6-10 Minutes	1698
3	01-Dec-21	Unknown / Data Quality	1277
4	01-Dec-21	16-20 Minutes	730

Skewness (numeric columns):

count_of_appointments   2.633315

dtype: float64

Kurtosis (numeric columns):

count_of_appointments   10.438427

dtype: float64

Outlier Counts (IQR method):

{'count_of_appointments': 9291}

Appendix C: NHS Geography Detail from Appointment Data

Note: Region codes have changed since 2021/22. Opted not to substitute codes with actual names unless they appear in the data (for example sub_icb_location_name: NHS North East and North Cumbria ICB – 00L).

E40000003

	icb_ons_code	sub_icb_location_ons_code	sub_icb_location_name
0	E54000027	E38000256	NHS North West London ICB - W2U3Z
1	E54000028	E38000240	NHS North Central London ICB - 93C
2	E54000029	E38000255	NHS North East London ICB - A3A8R
3	E54000030	E38000244	NHS South East London ICB - 72Q
4	E54000031	E38000245	NHS South West London ICB - 36L

E40000005

	icb_ons_code	sub_icb_location_ons_code	sub_icb_location_name
0	E54000032	E38000237	NHS Kent and Medway ICB - 91Q
1	E54000034	E38000252	NHS Frimley ICB - D4U1Y
2	E54000042	E38000137	NHS Hampshire and Isle Of Wight ICB - 10R
3	E54000042	E38000253	NHS Hampshire and Isle Of Wight ICB - D9Y0V
4	E54000044	E38000136	NHS Buckinghamshire Oxfordshire and Berkshire …
5	E54000044	E38000223	NHS Buckinghamshire Oxfordshire and Berkshire …
6	E54000044	E38000221	NHS Buckinghamshire Oxfordshire and Berkshire …
7	E54000052	E38000246	NHS Surrey Heartlands ICB - 92A
8	E54000053	E38000021	NHS Sussex ICB - 09D
9	E54000053	E38000248	NHS Sussex ICB - 70F
10	E54000053	E38000235	NHS Sussex ICB - 97R

E40000006

	icb_ons_code	sub_icb_location_ons_code	sub_icb_location_name
0	E54000036	E38000089	NHS Cornwall and The Isles Of Scilly ICB - 11N
1	E54000037	E38000230	NHS Devon ICB - 15N
2	E54000038	E38000150	NHS Somerset ICB - 11X
3	E54000039	E38000222	NHS Bristol North Somerset and South Glouceste...
4	E54000040	E38000231	NHS Bath and North East Somerset Swindon and W...
5	E54000041	E38000045	NHS Dorset ICB - 11J
6	E54000043	E38000062	NHS Gloucestershire ICB - 11M

E40000007

	icb_ons_code	sub_icb_location_ons_code	sub_icb_location_name
0	E54000022	E38000239	NHS Norfolk and Waveney ICB - 26A
1	E54000023	E38000086	NHS Suffolk and North East Essex ICB - 06L
2	E54000023	E38000117	NHS Suffolk and North East Essex ICB - 06T
3	E54000023	E38000204	NHS Suffolk and North East Essex ICB - 07K
4	E54000024	E38000249	NHS Bedfordshire Luton and Milton Keynes ICB -...
5	E54000025	E38000049	NHS Hertfordshire and West Essex ICB - 06K
6	E54000025	E38000079	NHS Hertfordshire and West Essex ICB - 06N
7	E54000025	E38000197	NHS Hertfordshire and West Essex ICB - 07H
8	E54000026	E38000106	NHS Mid and South Essex ICB - 06Q
9	E54000026	E38000185	NHS Mid and South Essex ICB - 07G
10	E54000026	E38000007	NHS Mid and South Essex ICB - 99E
11	E54000026	E38000030	NHS Mid and South Essex ICB - 99F
12	E54000026	E38000168	NHS Mid and South Essex ICB - 99G
13	E54000056	E38000260	NHS Cambridgeshire and Peterborough ICB - 06H

```
E40000010
```

	icb_ons_code	sub_icb_location_ons_code	sub_icb_location_name
0	E54000008	E38000161	NHS Cheshire and Merseyside ICB - 01T
1	E54000008	E38000208	NHS Cheshire and Merseyside ICB - 12F
2	E54000008	E38000194	NHS Cheshire and Merseyside ICB - 02E
3	E54000008	E38000172	NHS Cheshire and Merseyside ICB - 01X
4	E54000008	E38000170	NHS Cheshire and Merseyside ICB - 01V
5	E54000008	E38000233	NHS Cheshire and Merseyside ICB - 27D
6	E54000008	E38000091	NHS Cheshire and Merseyside ICB - 01J
7	E54000008	E38000068	NHS Cheshire and Merseyside ICB - 01F
8	E54000008	E38000101	NHS Cheshire and Merseyside ICB - 99A
9	E54000048	E38000014	NHS Lancashire and South Cumbria ICB - 00Q
10	E54000048	E38000050	NHS Lancashire and South Cumbria ICB - 01A
11	E54000048	E38000228	NHS Lancashire and South Cumbria ICB - 01K
12	E54000048	E38000015	NHS Lancashire and South Cumbria ICB - 00R
13	E54000048	E38000034	NHS Lancashire and South Cumbria ICB - 00X
14	E54000048	E38000226	NHS Lancashire and South Cumbria ICB - 02M
15	E54000048	E38000227	NHS Lancashire and South Cumbria ICB - 01E

	icb_ons_code	sub_icb_location_ons_code	sub_icb_location_name

16	E54000048	E38000200	NHS Lancashire and South Cumbria ICB - 02G
17	E54000057	E38000217	NHS Greater Manchester ICB - 14L
18	E54000057	E38000205	NHS Greater Manchester ICB - 02H
19	E54000057	E38000080	NHS Greater Manchester ICB - 01D
20	E54000057	E38000187	NHS Greater Manchester ICB - 02A
21	E54000057	E38000024	NHS Greater Manchester ICB - 00V
22	E54000057	E38000174	NHS Greater Manchester ICB - 01W
23	E54000057	E38000135	NHS Greater Manchester ICB - 00Y
24	E54000057	E38000143	NHS Greater Manchester ICB - 01G
25	E54000057	E38000016	NHS Greater Manchester ICB - 00T
26	E54000057	E38000263	NHS Greater Manchester ICB - 01Y

```
E40000011
```

	icb_ons_code	sub_icb_location_ons_code	sub_icb_location_name
0	E54000010	E38000175	NHS Staffordshire and Stoke-on-Trent ICB - 05W
1	E54000010	E38000173	NHS Staffordshire and Stoke-on-Trent ICB - 05V
2	E54000010	E38000153	NHS Staffordshire and Stoke-on-Trent ICB - 05Q
3	E54000010	E38000126	NHS Staffordshire and Stoke-on-Trent ICB - 05G
4	E54000010	E38000028	NHS Staffordshire and Stoke-on-Trent ICB - 04Y
5	E54000010	E38000053	NHS Staffordshire and Stoke-on-Trent ICB - 05D
6	E54000011	E38000257	NHS Shropshire Telford and Wrekin ICB - M2L0M
7	E54000013	E38000238	NHS Lincolnshire ICB - 71E
8	E54000015	E38000097	NHS Leicester Leicestershire and Rutland ICB -...
9	E54000015	E38000051	NHS Leicester Leicestershire and Rutland ICB -...
10	E54000015	E38000201	NHS Leicester Leicestershire and Rutland ICB -...
11	E54000018	E38000251	NHS Coventry and Warwickshire ICB - B2M3M
12	E54000019	E38000236	NHS Herefordshire and Worcestershire ICB - 18C
13	E54000055	E38000258	NHS Birmingham and Solihull ICB - 15E
14	E54000058	E38000261	NHS Derby and Derbyshire ICB - 15M
15	E54000059	E38000262	NHS Northamptonshire ICB - 78H
16	E54000060	E38000243	NHS Nottingham and Nottinghamshire ICB - 52R
17	E54000060	E38000008	NHS Nottingham and Nottinghamshire ICB - 02Q
18	E54000062	E38000259	NHS Black Country ICB - D2P2L

27

E40000012

	icb_ons_code	sub_icb_location_ons_code	sub_icb_location_name
0	E54000050	E38000130	NHS North East and North Cumbria ICB - 00L
1	E54000050	E38000163	NHS North East and North Cumbria ICB - 00N
2	E54000050	E38000176	NHS North East and North Cumbria ICB - 00P
3	E54000050	E38000215	NHS North East and North Cumbria ICB - 01H
4	E54000050	E38000234	NHS North East and North Cumbria ICB - 84H
5	E54000050	E38000127	NHS North East and North Cumbria ICB - 99C
6	E54000050	E38000247	NHS North East and North Cumbria ICB - 16C
7	E54000050	E38000212	NHS North East and North Cumbria ICB - 13T
8	E54000051	E38000241	NHS Humber and North Yorkshire ICB - 42D
9	E54000051	E38000052	NHS Humber and North Yorkshire ICB - 02Y
10	E54000051	E38000085	NHS Humber and North Yorkshire ICB - 03F
11	E54000051	E38000119	NHS Humber and North Yorkshire ICB - 03H
12	E54000051	E38000122	NHS Humber and North Yorkshire ICB - 03K
13	E54000051	E38000188	NHS Humber and North Yorkshire ICB - 03Q
14	E54000054	E38000232	NHS West Yorkshire ICB - 36J
15	E54000054	E38000225	NHS West Yorkshire ICB - 15F
16	E54000054	E38000254	NHS West Yorkshire ICB - X2C4Y
17	E54000054	E38000025	NHS West Yorkshire ICB - 02T
18	E54000054	E38000190	NHS West Yorkshire ICB - 03R
19	E54000061	E38000146	NHS South Yorkshire ICB - 03N
20	E54000061	E38000044	NHS South Yorkshire ICB - 02X
21	E54000061	E38000006	NHS South Yorkshire ICB - 02P
22	E54000061	E38000141	NHS South Yorkshire ICB - 03L

	icb_ons_code	sub_icb_location_ons_code	sub_icb_location_name

Appendix D: Appointment Totals by Service Setting

	service_setting	count_of_appointments
0	Extended Access Provision	2176807
1	General Practice	270811691
2	Other	5420076
3	Primary Care Network	6557386
4	Unmapped	11080810

Appendix E: Appointment Statistics

```
appointment_summary.describe()
```

	total_appointments
count	3.000000e+01
mean	2.476015e+07
std	3.472766e+06
min	1.600788e+07
25%	2.351528e+07
50%	2.510119e+07
75%	2.709552e+07
max	3.040507e+07

```
record_count.head(10)
```

	appointment_month	size
0	2020-01	20889
1	2020-02	20689
2	2020-03	21350
3	2020-04	19124
4	2020-05	18338
5	2020-06	18844
6	2020-07	19502
7	2020-08	19247
8	2020-09	20043
9	2020-10	20122

```
record_count.describe()
```

	size
count	30.000000
mean	19894.033333
std	679.651203
min	18338.000000
25%	19386.500000
50%	19971.000000
75%	20363.750000
max	21350.000000

Appendix F: Monthly Totals in nc

nc_monthly_totals

	appointment_month	count_of_appointments
0	2021-08	23852171
1	2021-09	28522501
2	2021-10	30303834
3	2021-11	30405070
4	2021-12	25140776
5	2022-01	25635474
6	2022-02	25355260
7	2022-03	29595038
8	2022-04	23913060
9	2022-05	27495508
10	2022-06	25828078

Appendix G: Seasonal Totals in nc

nc_season_totals

	season	count_of_appointments
0	Autumn	89231405
1	Spring	81003606
2	Summer	49680249
3	Winter	76131510

Appendix H: Seasonal Totals in ar

ar_season_totals

	season	count_of_appointments
0	Autumn	169309194
1	Spring	212095918
2	Summer	165536412
3	Winter	195863001

Appendix I: Social Media Analysis

Evidence of generic healthcare terms in #hashtag analysis. These metrics were generated outside the Notebook submitted with the project deliverables.
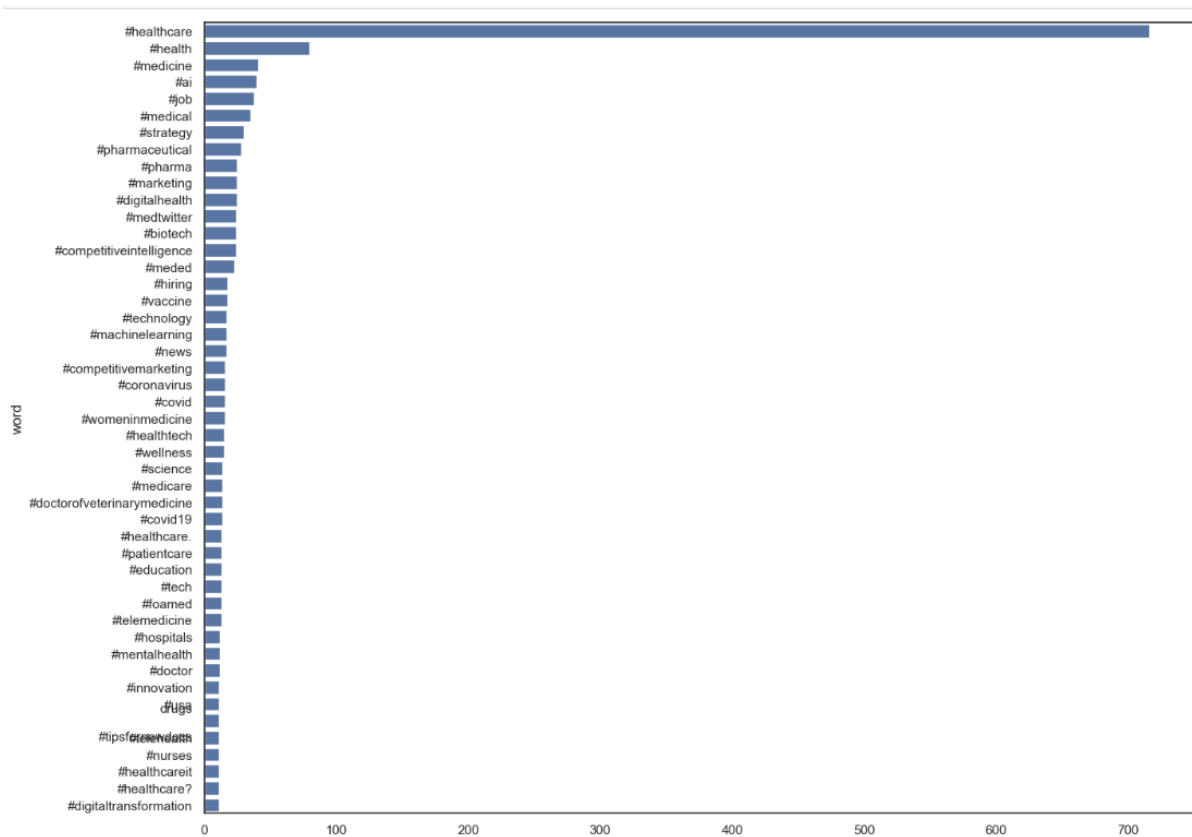
The text labels in the above barplot are very small so top 10 hashtags from the tweets data set printed below;

	word	count
0	#healthcare	716
1	#health	80
2	#medicine	41
3	#ai	40
4	#job	38
5	#medical	35
6	#strategy	30
7	#pharmaceutical	28
8	#pharma	25
9	#marketing	25

Finally, the statistical summary of the tweets data set indicate particularly low numbers of favourites and retweets;

```
tweets.describe()
```

	tweet_id	tweet_retweet_count	tweet_favorite_count
count	1.174000e+03	1174.000000	1174.00000
mean	1.567612e+18	8.629472	0.37138
std	2.427553e+13	29.784675	2.04470
min	1.567574e+18	0.000000	0.00000
25%	1.567590e+18	0.000000	0.00000
50%	1.567611e+18	1.000000	0.00000
75%	1.567633e+18	3.000000	0.00000
max	1.567655e+18	303.000000	42.00000

Appendix J: Appointment Status Totals – ar

```
status_totals
```

	appointment_status	count_of_appointments
0	Attended	677755876
1	DNA	30911233
2	Unknown	34137416

91% attended 4% unkown 5% DNA - further exploration of other columns req'd to find causes for non attendance