

# CAPACITY ALLOCATION AND UTILISATION OF RESOURCES IN THE NHS

Kelvin Dhondee

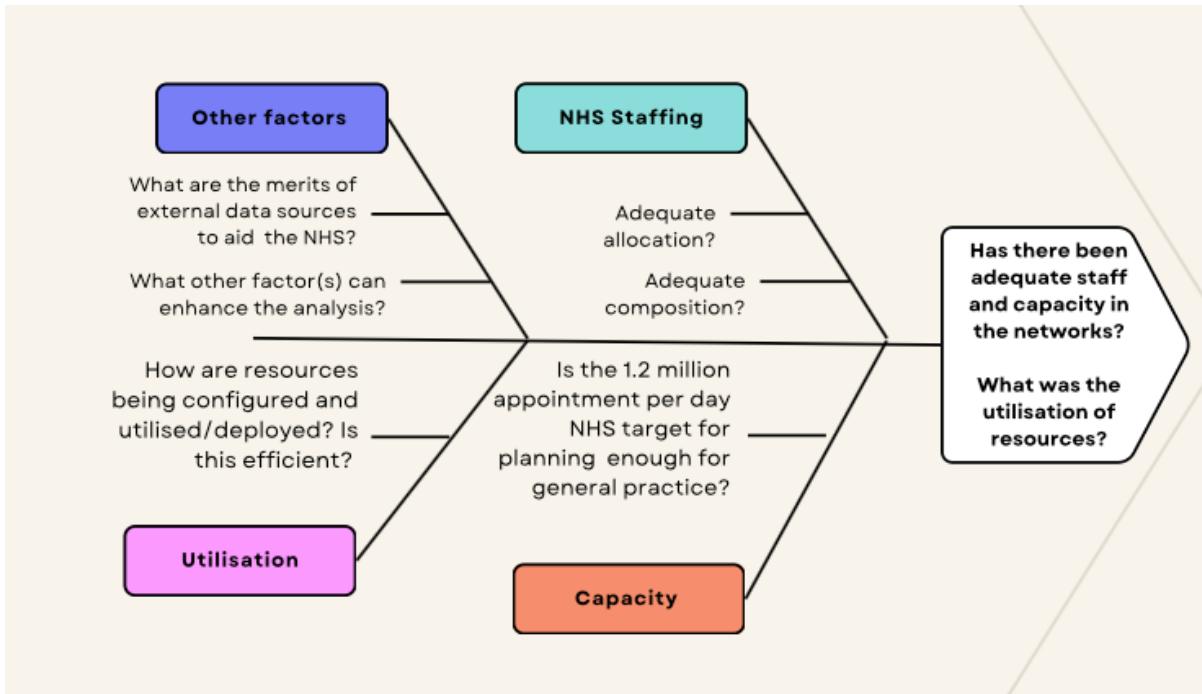
## Table of Contents

<b><u>BACKGROUND INFORMATION.....</u></b>	<b>2</b>
<b><u>ANALYTICAL APPROACH.....</u></b>	<b>3</b>
<b>SCOPE OF ANALYSIS.....</b>	<b>3</b>
<b>KEY INFORMATION IN RELATION TO THE ICB-LEVEL ANALYSIS .....</b>	<b>4</b>
<b>DATA PROVIDED FOR THE ANALYSIS .....</b>	<b>4</b>
<b>LIMITATIONS OF THE DATA .....</b>	<b>4</b>
<b>PYTHON LIBRARIES UTILISED .....</b>	<b>5</b>
<b>DATA WRANGLING AND VALIDATION .....</b>	<b>5</b>
<b>FILES MERGED WITH THE ORIGINAL DATA .....</b>	<b>6</b>
<b>REMOVAL OF DUPLICATE VALUES (FROM AR DATASET).....</b>	<b>7</b>
<b>OUTLIER ANALYSIS .....</b>	<b>7</b>
<b><u>VISUALISATIONS .....</u></b>	<b>9</b>
<b>SELECTED EXPLORATORY VISUALISATIONS GENERATED IN THE EDA STAGE OF THE ANALYSIS. ....</b>	<b>9</b>
<b>SELECTED EXPLANATORY VISUALS .....</b>	<b>11</b>
<b><u>PATTERNS AND INSIGHTS IDENTIFIED .....</u></b>	<b>16</b>
<b><u>RECOMMENDATIONS .....</u></b>	<b>22</b>
<b><u>FURTHER ACTIONS .....</u></b>	<b>23</b>
<b><u>APPENDICES.....</u></b>	<b>24</b>
<b>APPENDIX A.....</b>	<b>24</b>
KEY TERMS .....	24
<b>APPENDIX B.....</b>	<b>24</b>
METRICS GENERATED AND/OR UTILISED .....	24
<b>APPENDIX C .....</b>	<b>25</b>
LIST OF KEY ANALYSIS PERFORMED .....	25

## Background information

This analysis was commissioned by National Health Service (NHS) England. It addresses the following two key questions:

- 1) Has there been adequate staff and capacity in the networks?
- 2) What was the utilisation of resources?



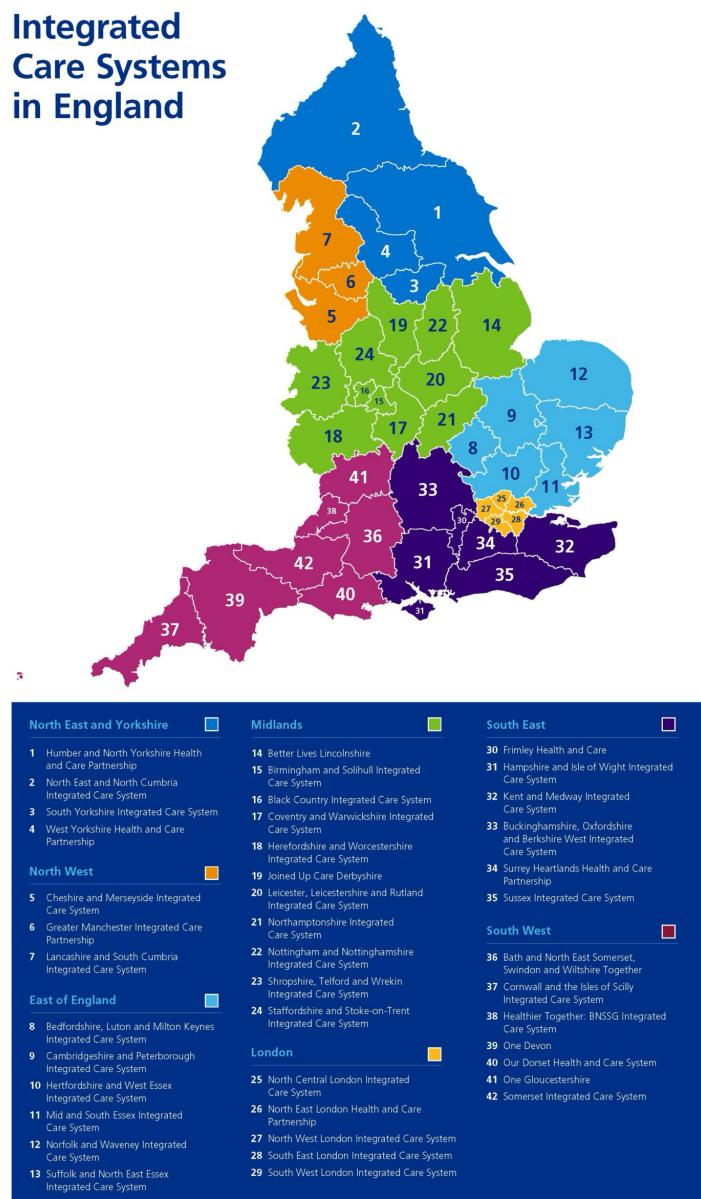
The analysis was categorised into the following themes as per the above Fishbone diagram:

- a) Capacity provision – Is the 1.2 million daily appointments target for general practice, used by the NHS for planning purposes still adequate?
- b) Capacity utilisation – How were resources utilised and deployed? Was this done in an efficient manner? (To minimise waste and maximise patient health outcomes).
- c) Staffing in general practice – Adequate allocation and composition?
- d) Other factors – Twitter data and deprivation levels.

# Analytical approach

## Scope of analysis

Owing to time constraints, a decision was made to focus the analysis at the ICB level. For the purpose of conducting an effective analysis, was deemed to represent a balance between not being too generic (7 NHS regions) and not being too localised (106 sub-ICB locations).



## Key information in relation to the ICB-level Analysis

Integrated care boards (ICBs) replaced clinical commissioning groups (CCGs) in the NHS in England from 1 July 2022:

A way was identified to convert CCG codes into ICB codes and in turn compile the information in relation to registered patient list sizes and workforce data for periods prior to June 2022: <https://www.sbs.nhs.uk/supplier-information/ccg-icb-list/>.

However, owing to time constraints, a decision was made to only focus on June 2022 appointments as the related list size (01.07.2022) and workforce data (30.06.2022) was readily available for easy merging and speedy analysis.

<https://digital.nhs.uk/data-and-information/publications/statistical/patients-registered-at-a-gp-practice/july-2022>

<https://digital.nhs.uk/data-and-information/publications/statistical/general-and-personal-medical-services/30-june-2022>

## Data provided for the analysis

File	Description
actual_duration.csv	Sub-ICB level data with related ICB ONS codes, Region ONS Codes, Actual Duration, Date
appointmentsRegional.csv	ICB level aggregation, Appointment Status, HCP Types, Appointment Modes, TBBA, Month
national_categories.xlsx	Sub-ICB level data with related ICB ONS codes, Service Settings, Context Types, National Categories, Date/Month
tweets.csv	Twitter data relating to healthcare
metadata_nhs.txt	Please refer to this for more information. See link in previous section

The original datasets were not amended for this analysis.

## Limitations of the data

- The raw data is drawn from an experimental statistics publication.
- The count of appointments is an estimation (not 100% Practice Coverage). See metadata.
- Data in the appointments in general practice collection contains number of data quality issues due to variation in working methods and recording between different practices and PCNs. See link below for more information:  
<https://digital.nhs.uk/data-and-information/publications/statistical/appointments-in-general-practice/appointments-in-general-practice-supporting-information#data-quality>

# Python libraries utilised

## 2. Pre-analysis configuration and importing the data

### 2.1 Standard library imports

```
[1]: # Being the import of necessary libraries.

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import nltk #!pip install vader_lexicon
nltk.download('vader_lexicon')
from nltk.sentiment import SentimentIntensityAnalyzer
import matplotlib.dates as mdates
from matplotlib.ticker import FuncFormatter
from datetime import date, time, datetime
from statsmodels.tsa.seasonal import seasonal_decompose
import warnings
warnings.filterwarnings('ignore')

[nltk_data] Downloading package vader_lexicon to
[nltk_data]   /Users/kelvin/nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
```

# Data wrangling and validation

- Common functions were utilised in order to validate data and generate standard function were set up see – 2.2 of ipynb file.
- The validate\_data() function conveniently displays the data, performs standard checks (nulls, unique values, duplicates, data types,) and provides summary statistics. This was carried out for all three data sets.

```
# Function to validate data.

def validate_data(df):
    # view DataFrame.
    print("\nFirst 5 rows of DataFrame:")
    display(df.head())

    print("\nLast 5 rows of Dataframe:")
    display(df.tail())

    shape = df.shape
    print("\nShape of DataFrame:")
    print(shape)

    # review metadata.
    print("\nDataFrame Info:")
    info = df.info()

    # check for missing values.
    missing_values = df.isnull().sum()
    print("\nNumber of nulls per column:")
    print(missing_values)

    # check for unique values.
    unique_counts = df.nunique()
    print("\nUnique values per column:")
    print(unique_counts)

    # check for duplicates.
    duplicate_counts = df.duplicated().sum()
    print("\nNumber of duplicate rows:")
    print(duplicate_counts)

    # summary statistics.
    summary_stats = df.describe().round(2)
    print("\nSummary statistics:")
    print(summary_stats)
```

- Having perused the metadata and conducted initial research, the sub- ICB ONS and location codes in the ad file were not deemed necessary for the analysis .
- Columns were rearranged accordingly post-merge with the icb\_region\_name.csv (see below).

## Files merged with the original data

It was not possible to merge the original data files on Python in any meaningful way.

This is due to overlap in the time coverage and the data being aggregated at different levels.

```
[53]: # Determine the minimum and maximum dates in the ad DataFrame.  
# Use appropriate docstrings.  
  
ad_date_min = (ad['appointment_date'].dt.date.min())  
ad_date_max = (ad['appointment_date'].dt.date.max())  
  
print(f"The minimum date in the ad DataFrame is {ad_date_min}")  
print(f"The maximum date in the ad DataFrame is {ad_date_max}")  
  
The minimum date in the ad DataFrame is 2021-12-01  
The maximum date in the ad DataFrame is 2022-06-30  
  
[54]: # Determine the minimum and maximum dates in the nc DataFrame.  
# Use appropriate docstrings.  
  
nc_date_min = (nc['appointment_date'].dt.date.min())  
nc_date_max = (nc['appointment_date'].dt.date.max())  
  
print(f"The minimum date in the nc DataFrame is {nc_date_min}")  
print(f"The maximum date in the nc DataFrame is {nc_date_max}")  
  
The minimum date in the nc DataFrame is 2021-08-01  
The maximum date in the nc DataFrame is 2022-06-30  
  
[55]: # Determine the minimum and maximum dates in the ar DataFrame.  
# Use appropriate docstrings.  
  
ar_date_min = (ar['appointment_month'].min())  
ar_date_max = (ar['appointment_month'].max())  
  
print(f"The minimum date in the ar DataFrame is {ar_date_min}")  
print(f"The maximum date in the ar DataFrame is {ar_date_max}")  
  
The minimum date in the ar DataFrame is 2020-01-01 00:00:00  
The maximum date in the ar DataFrame is 2022-06-01 00:00:00  
  
**The DataFrames cover different time periods and there are overlaps.  
As such, there may also be overlap in terms of the count of appointments featured in the datasets.  
This is a key consideration and also given that the data are aggregated at different levels, it does not make sense to merge the  
DataFrames at all.**
```

However, the analysis was enriched with the merge of the following files (sections are that of the ipynb file):

- icb\_region\_names.csv (section 2.3) To access NHS region names.
- reg\_list\_size.csv (section 9.1) To access Total NHS Registered Patient List Size Jan 2020 – Jul 2022
- imd\_score.csv (section 9.4 onwards -multiple merges) To access IMD Mean Score and Standard Deviation for all ICBs
- icb\_patient\_staffing.csv (section 10.1.1 onwards – multiple merges) To access GP Workforce data 30.06.2022 and Registered Practice List Size 01.07.2022 per ICB.

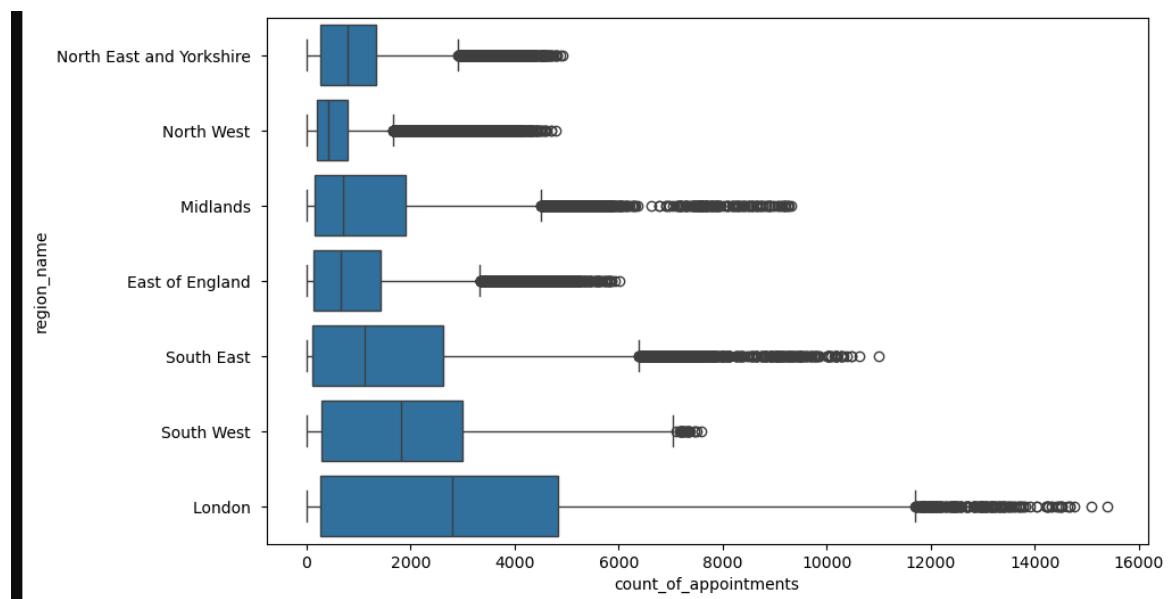
## Removal of duplicate values (from ar dataset)

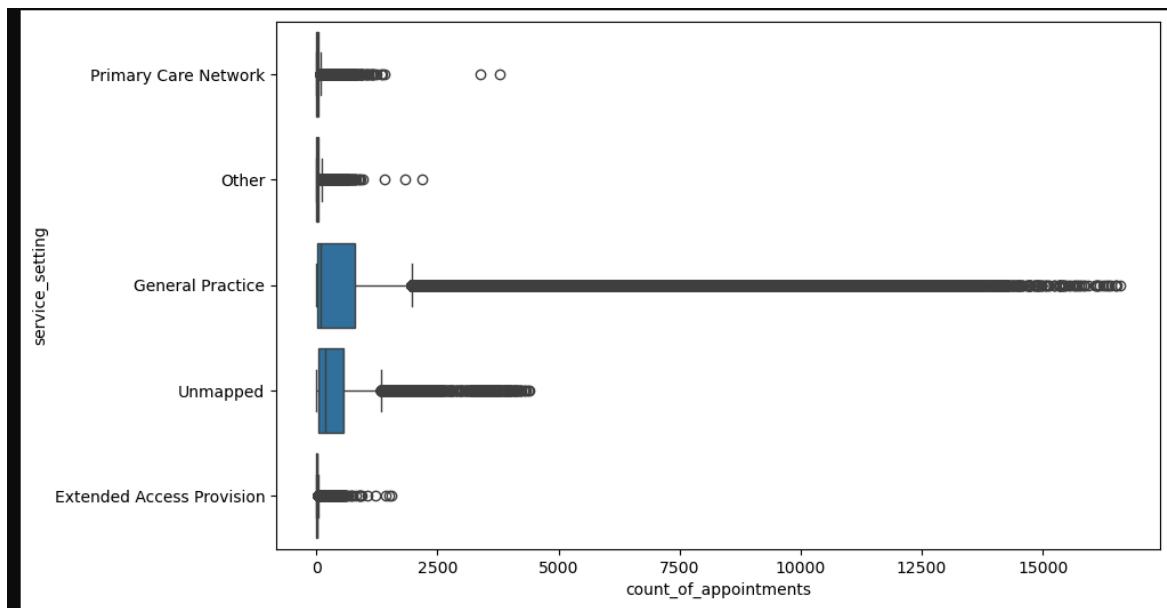
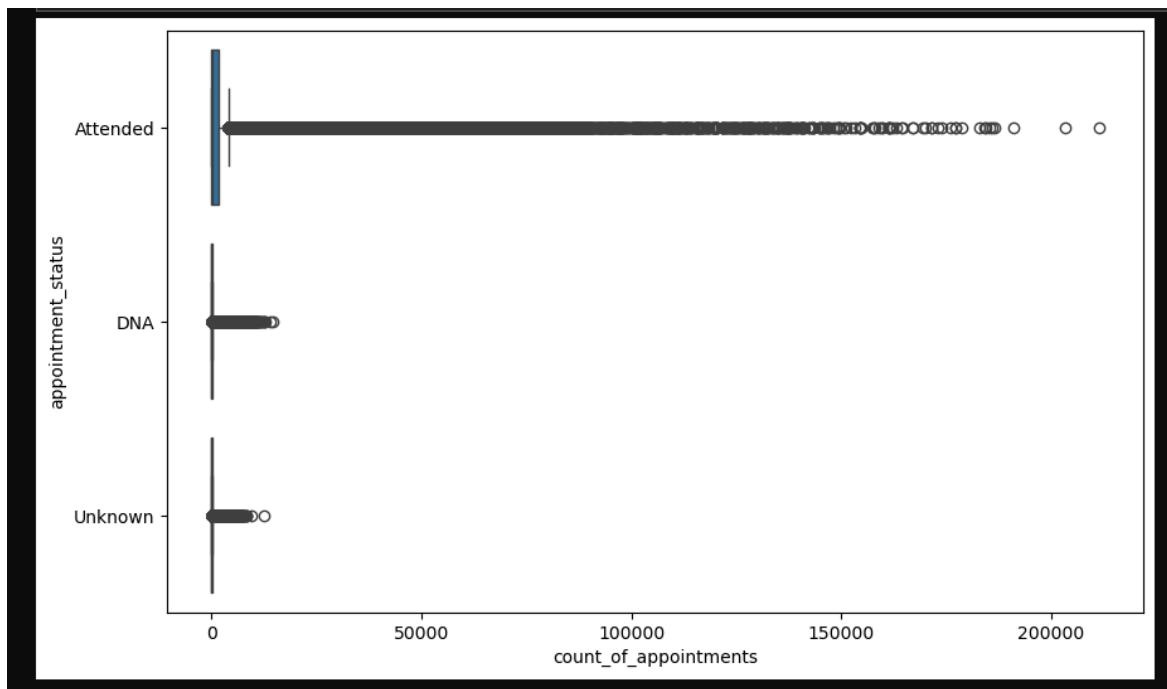
The ar dataset does not feature the related sub-ICB columns. It is unusual for an aggregated dataset to have duplicates and a possible explanation for why this has happened is the removal of the related sub-ICB columns after aggregation at ICB/regional level

It is not possible to ascertain whether these are actual duplicates. However, given that it only represents 3.62% of the data, these will be removed.

## Outlier Analysis

The numerical variable (count\_of\_appointments) in all three datasets was visualised with box plots.





The data has the characteristic of a high number of outliers, with a high count of lower number of appointment entries.

*Possible explanations:*

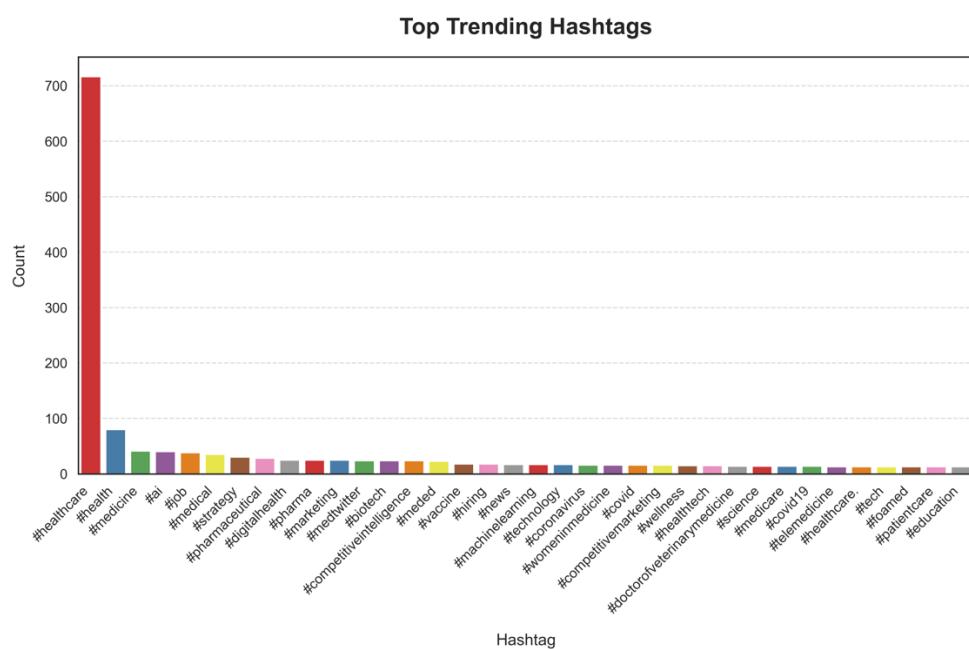
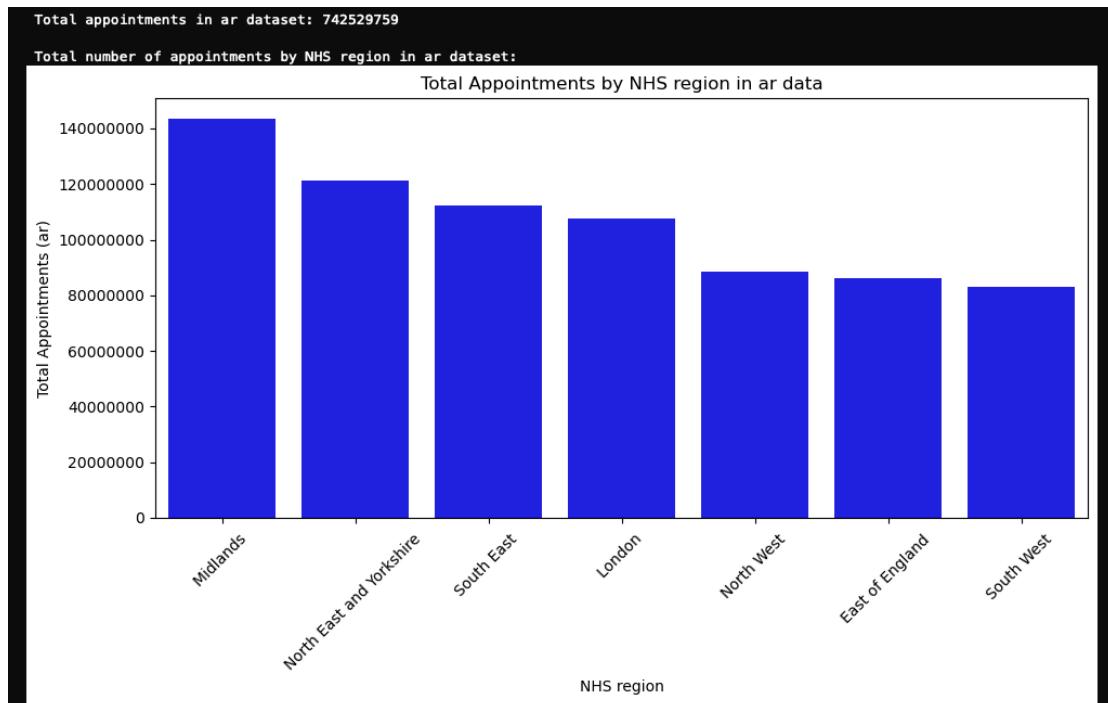
1. The difference between the high number of appointments of weekdays compared to weekends, seasonal variations, pandemic considerations etc.
2. Regional variations: urban v rural considerations.

As such, no further action was taken. Options could have included the removal of outlier values, min-max scaling or z-score normalisation.

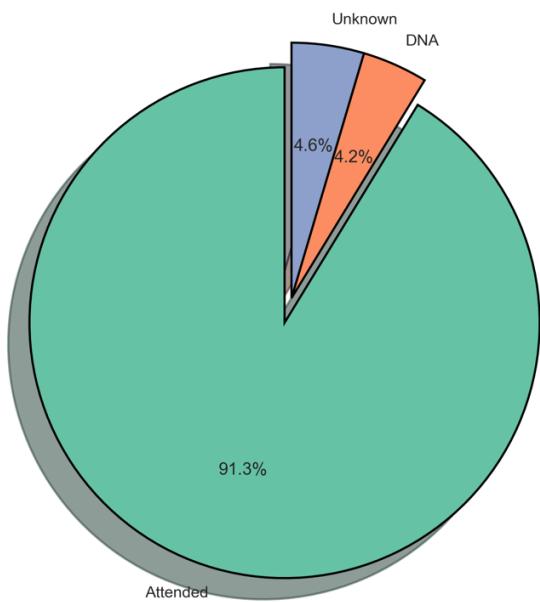
# Visualisations

Selected exploratory visualisations generated in the EDA stage of the analysis.

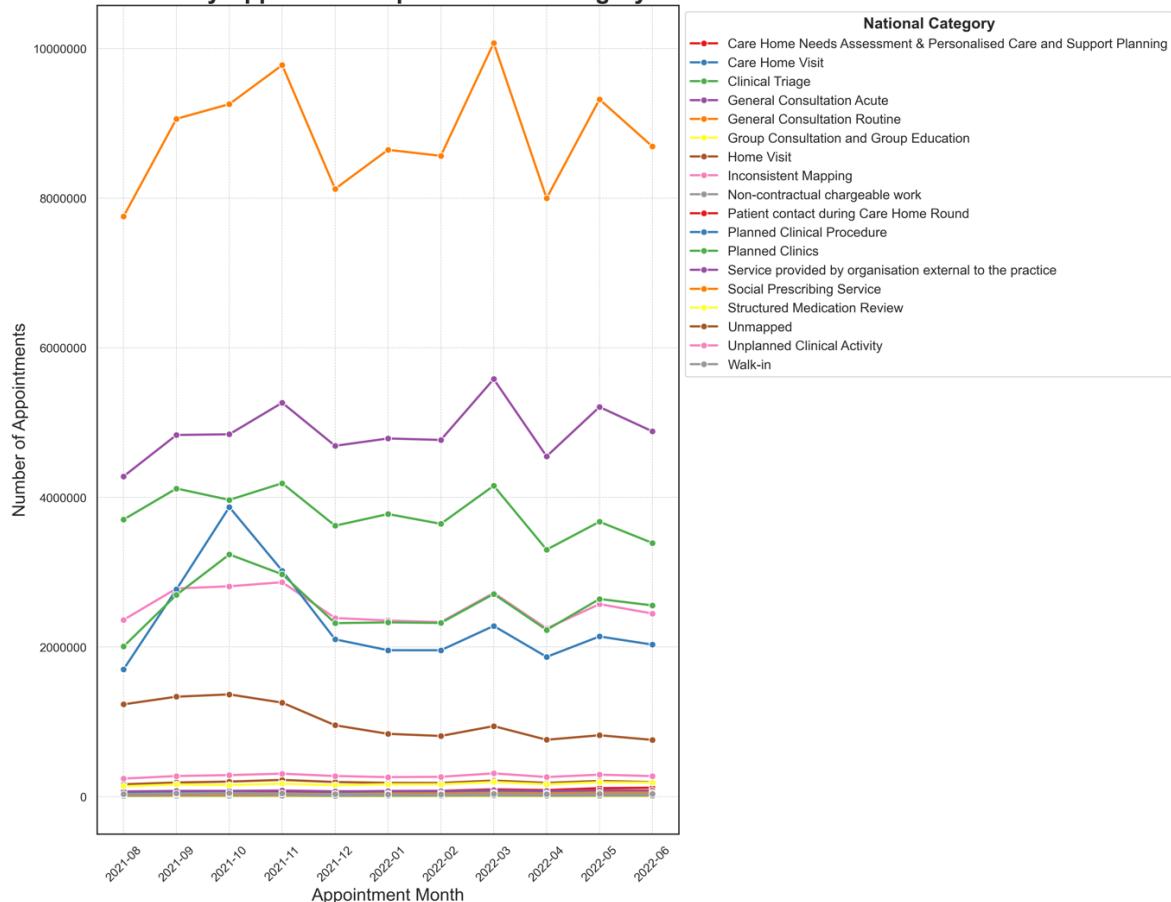
These aided familiarisation with the data.



Appointment Status % Jan 2020 - Jun 2022

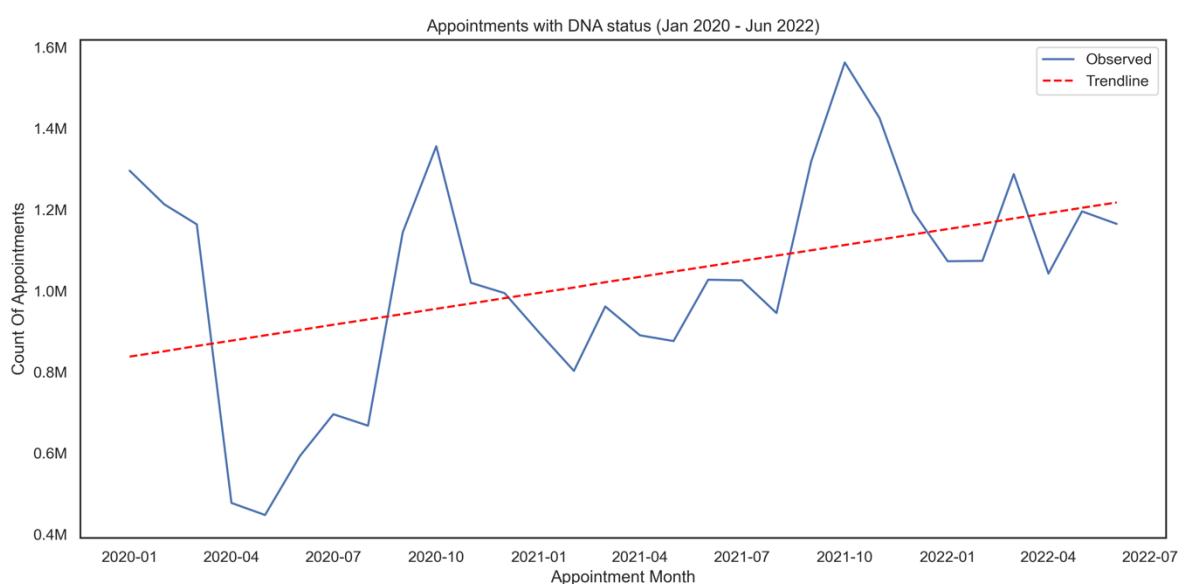
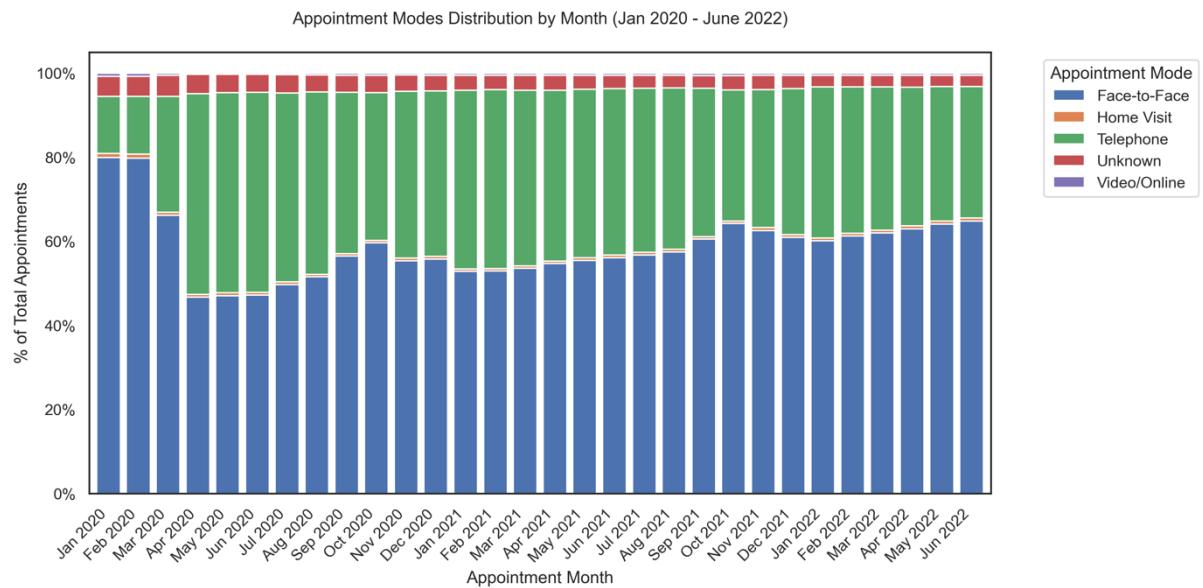


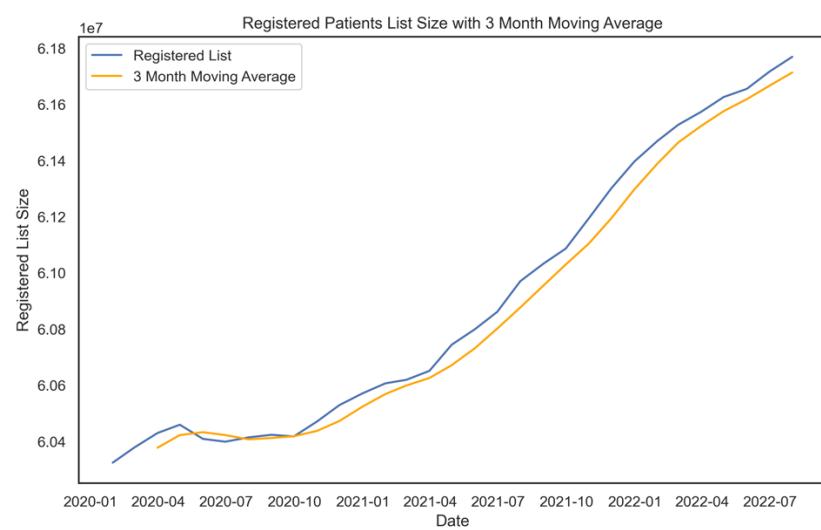
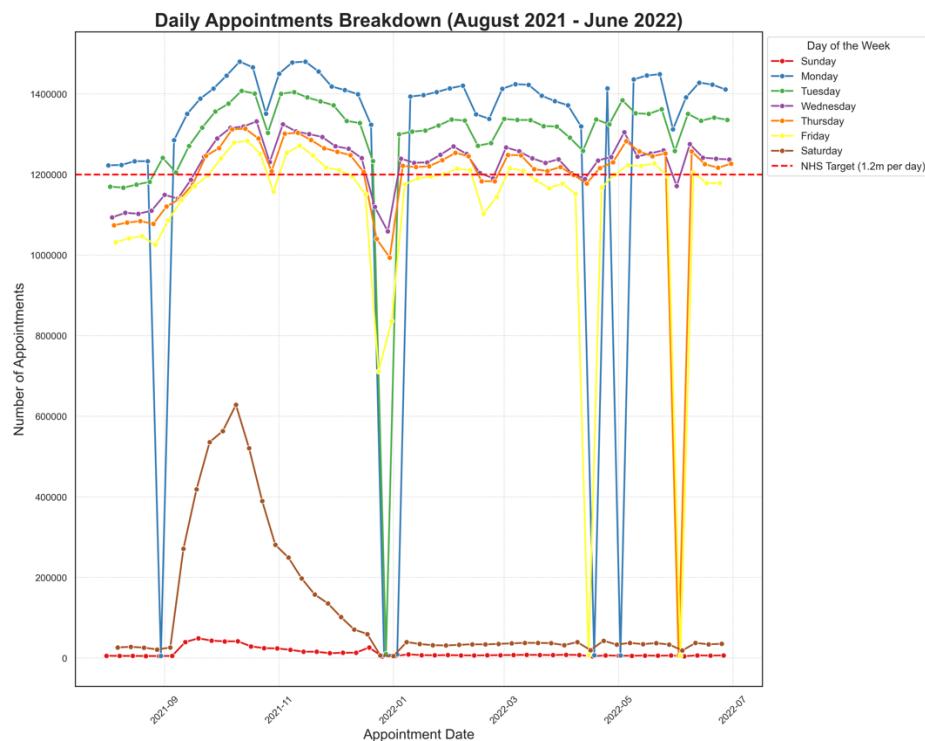
Monthly Appointments per National Category

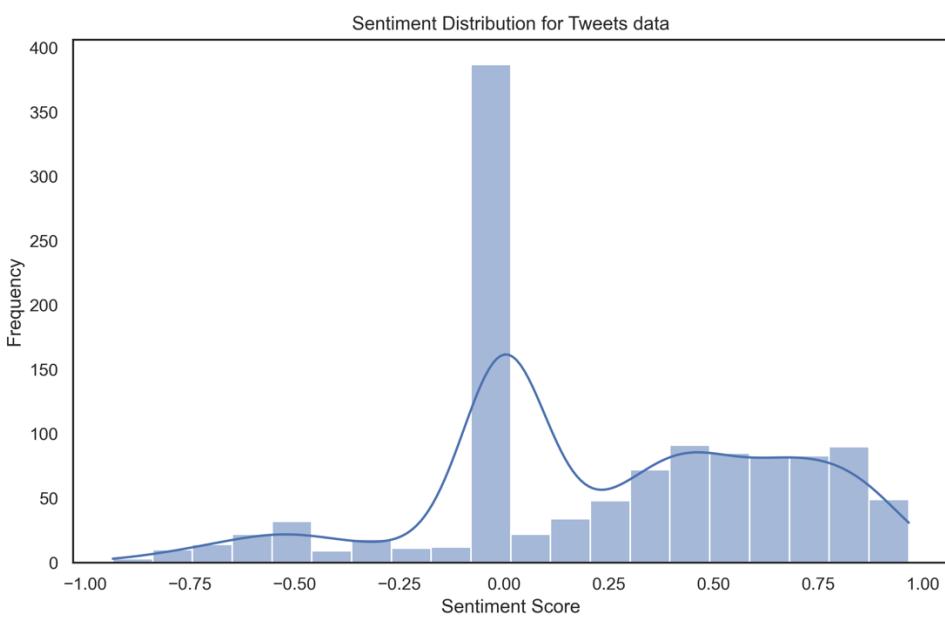
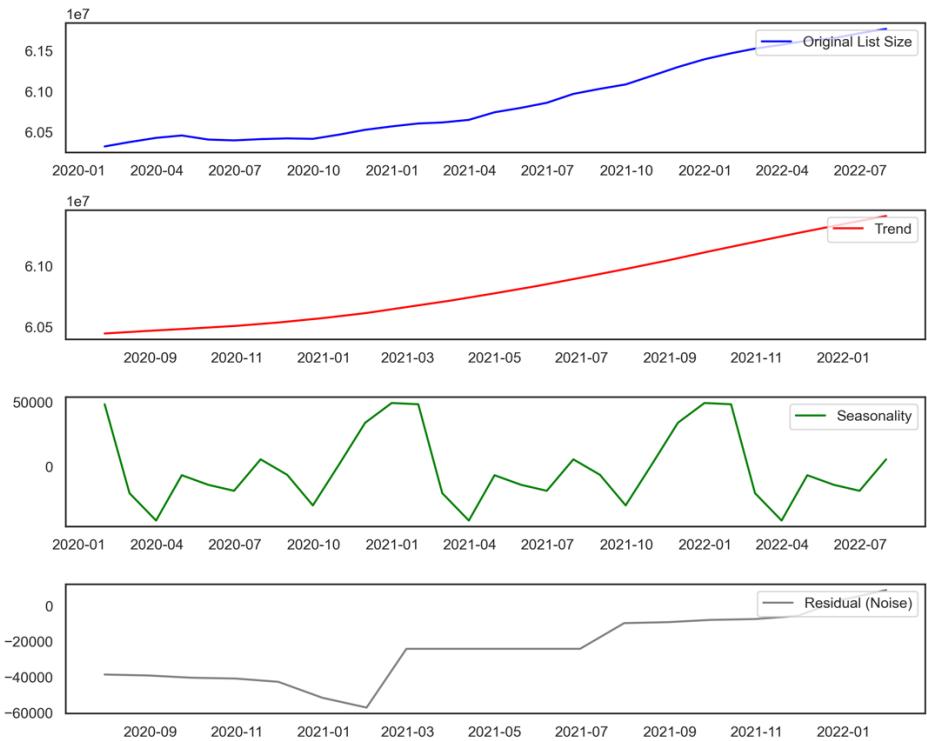


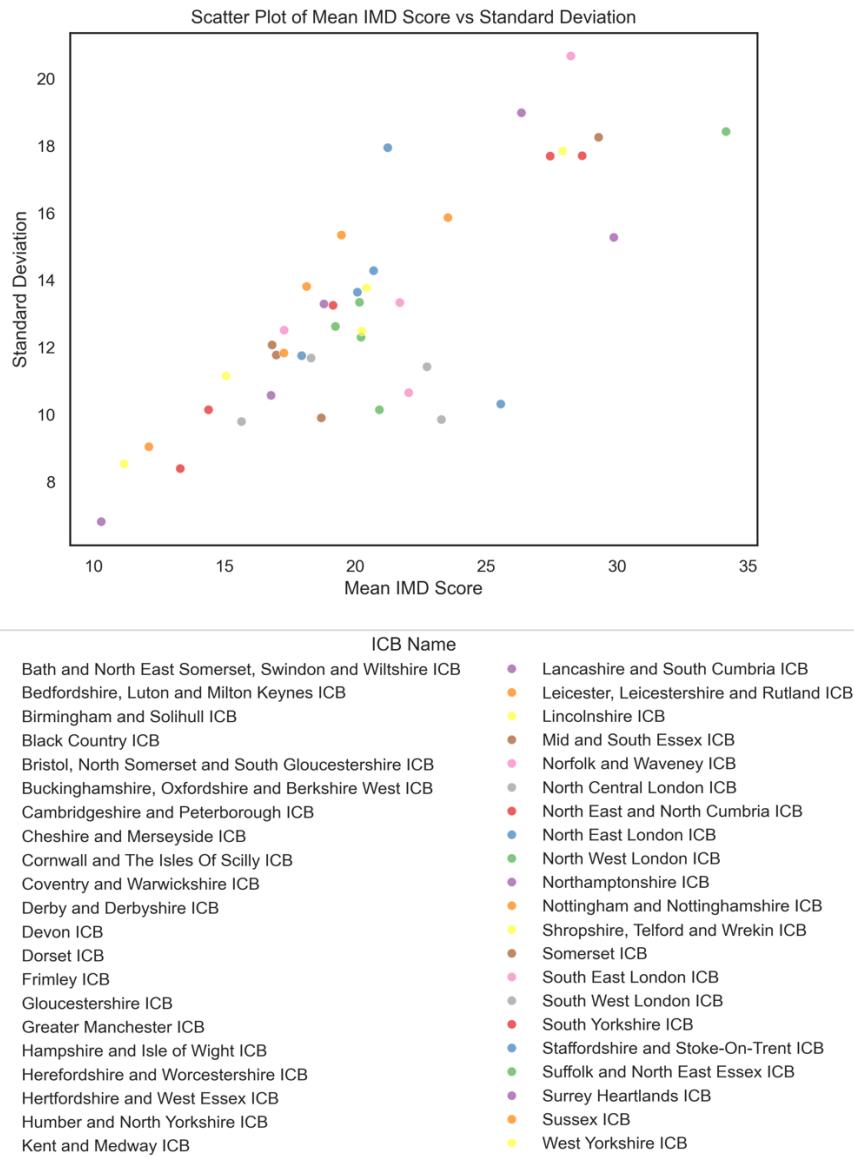
## Selected explanatory visuals

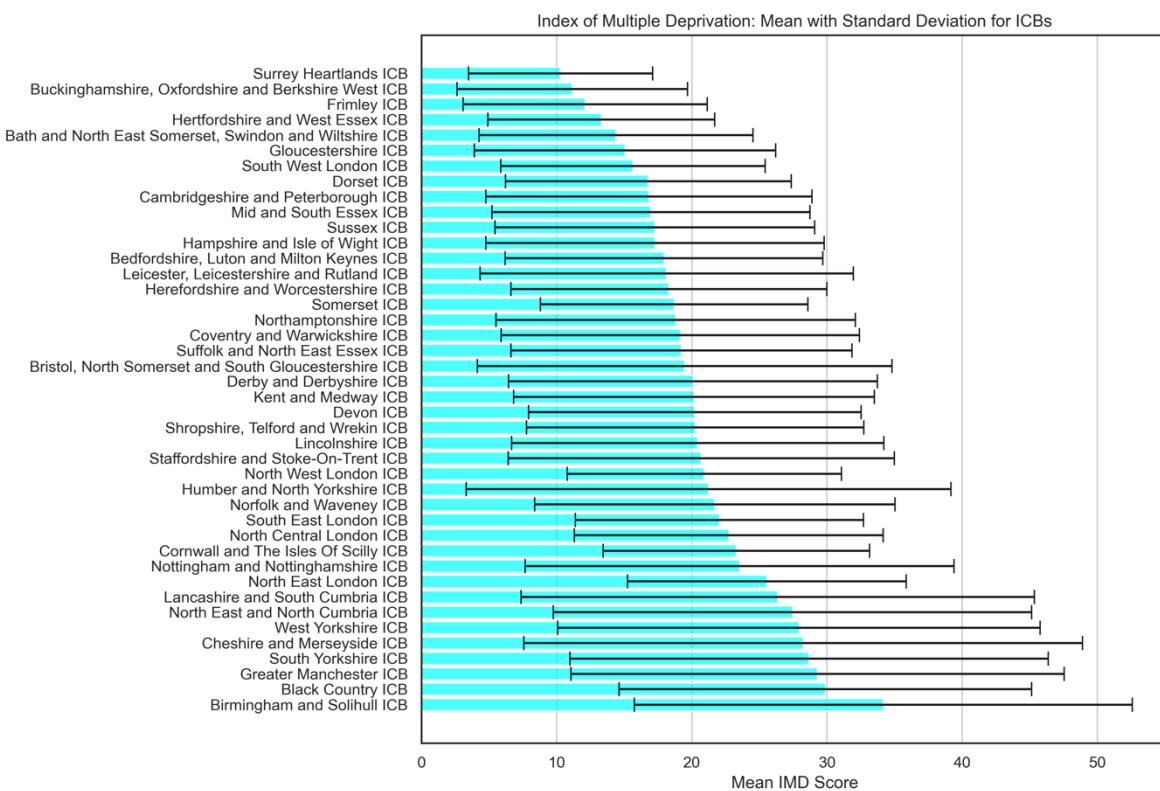
These could be adapted (e.g. catchy title, annotations, etc.) before presentation to stakeholders. Plot type chosen where trends and patterns are apparent in order to better communicate insights.





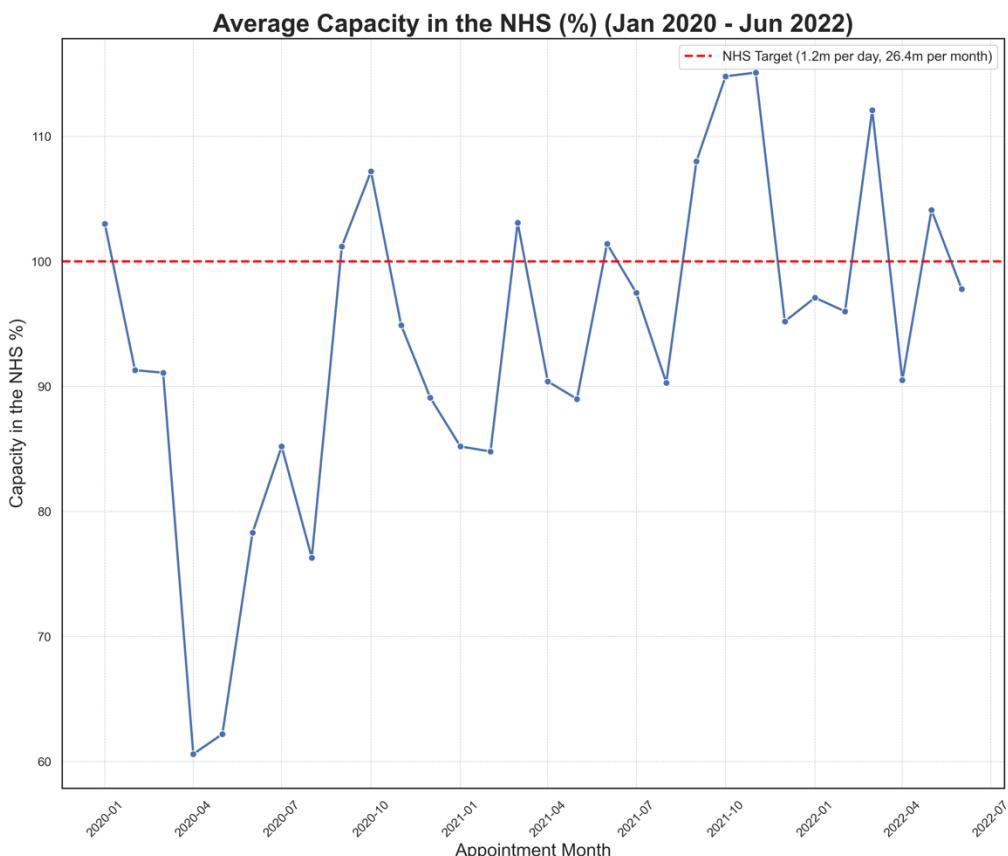
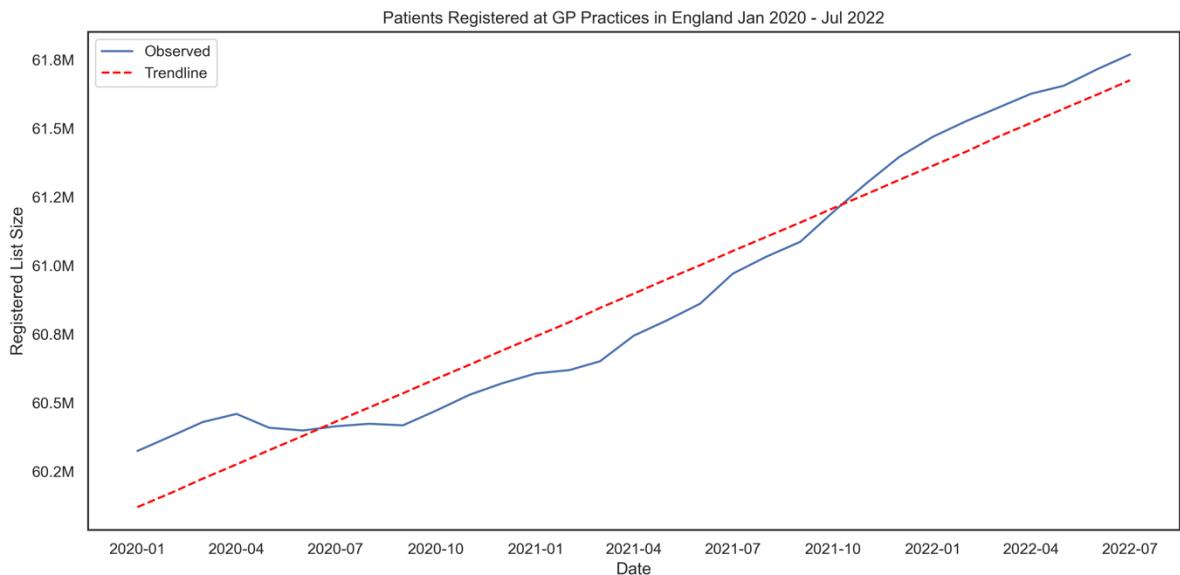






## Patterns and insights identified

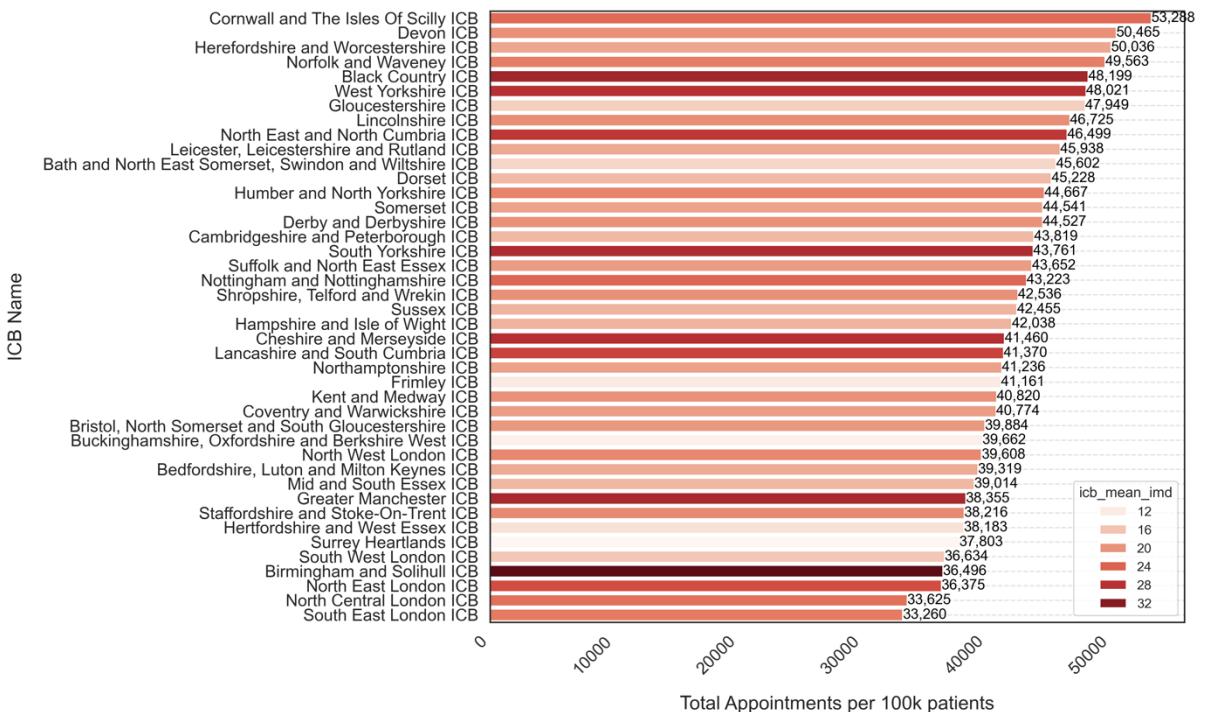
1. Patient list size growth on an upward trend (0.0765% per month). 1.2 million target appointments per day already inadequate capacity.



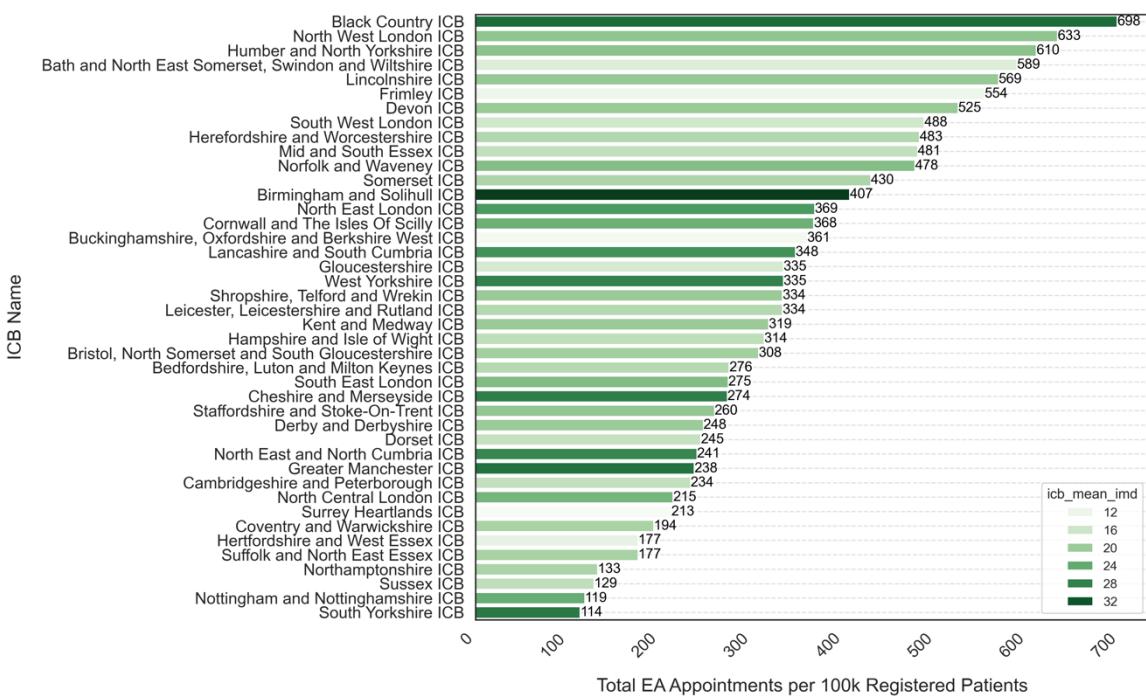
2. Deprivation levels matter for healthcare policy. There are disparities in access to healthcare in ICBs with high levels of deprivation. Capacity allocation not

adequate for ICBs with high deprivation when look at appointments per 100k patients.

**June 2022 Appointments per 100k registered patients**



**June 2022 Total EA Appointments per 100k Registered Patients**

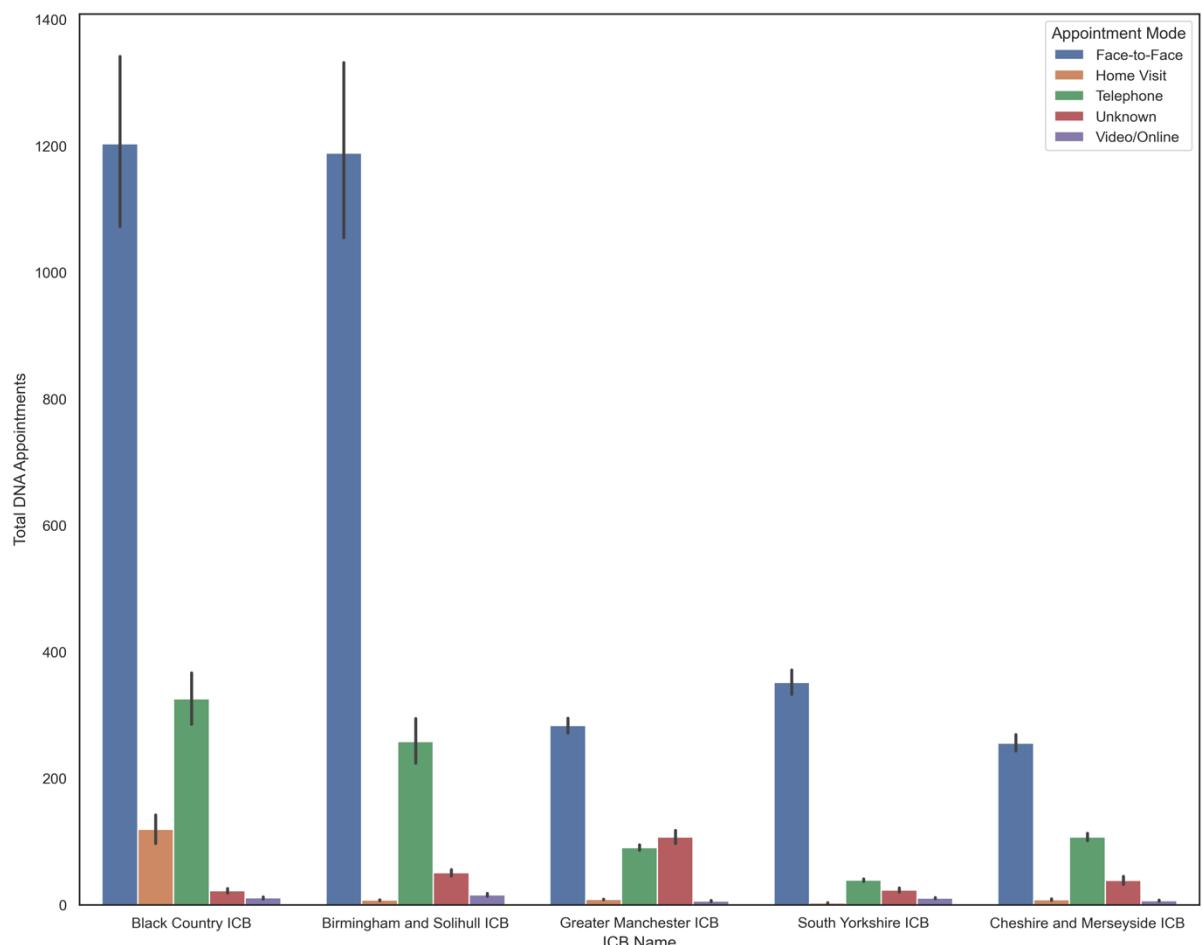


<https://www.health.org.uk/reports-and-analysis/briefings/integrated-care-systems-what-do-they-look-like>

<https://www.heec.co.uk/resource/understanding-the-index-of-multiple-deprivation-imd-in-public-health-research/>

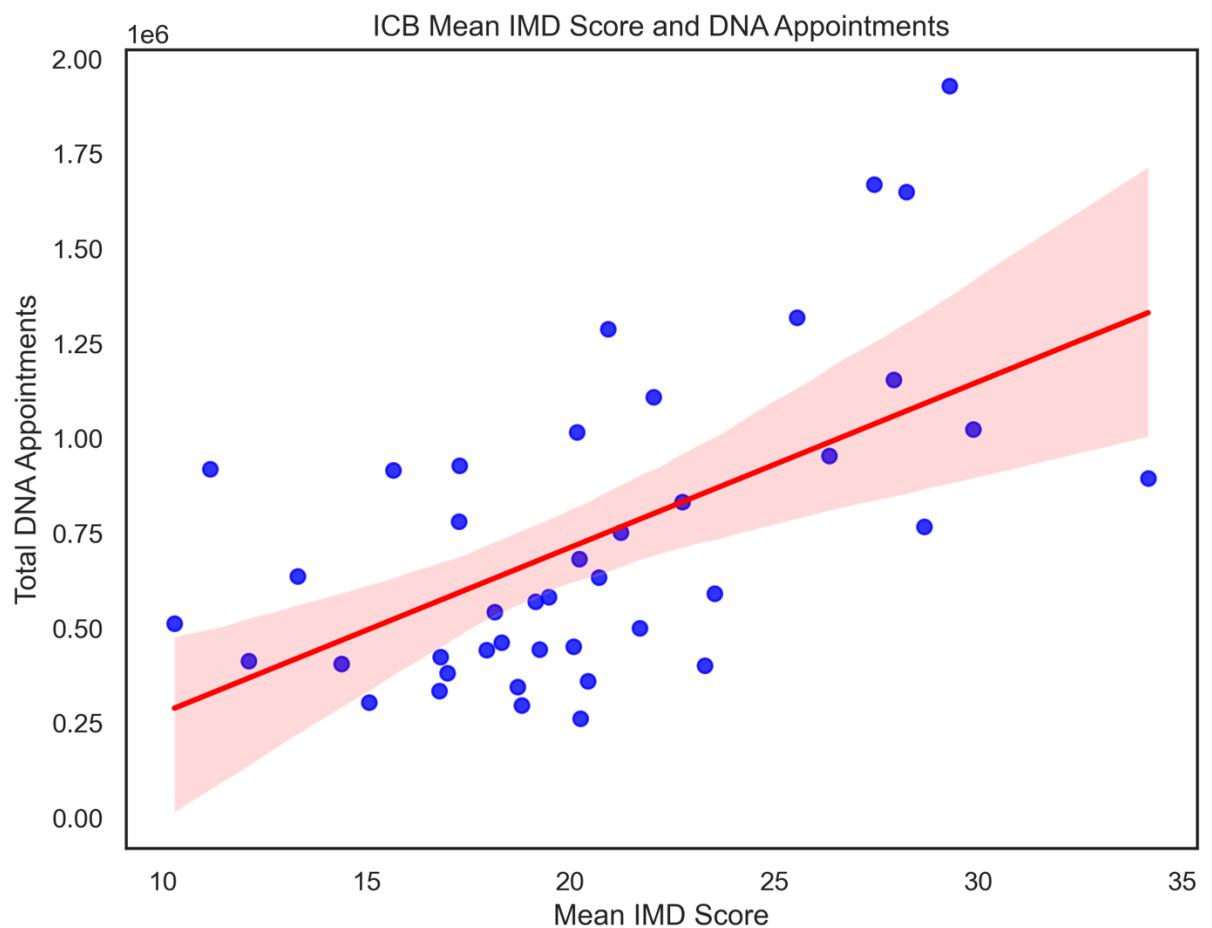
<https://www.heec.co.uk/component-library/data/>

3. Telephone appointments now accepted appointment delivery mode. Reduction may be explained by patient behaviour.

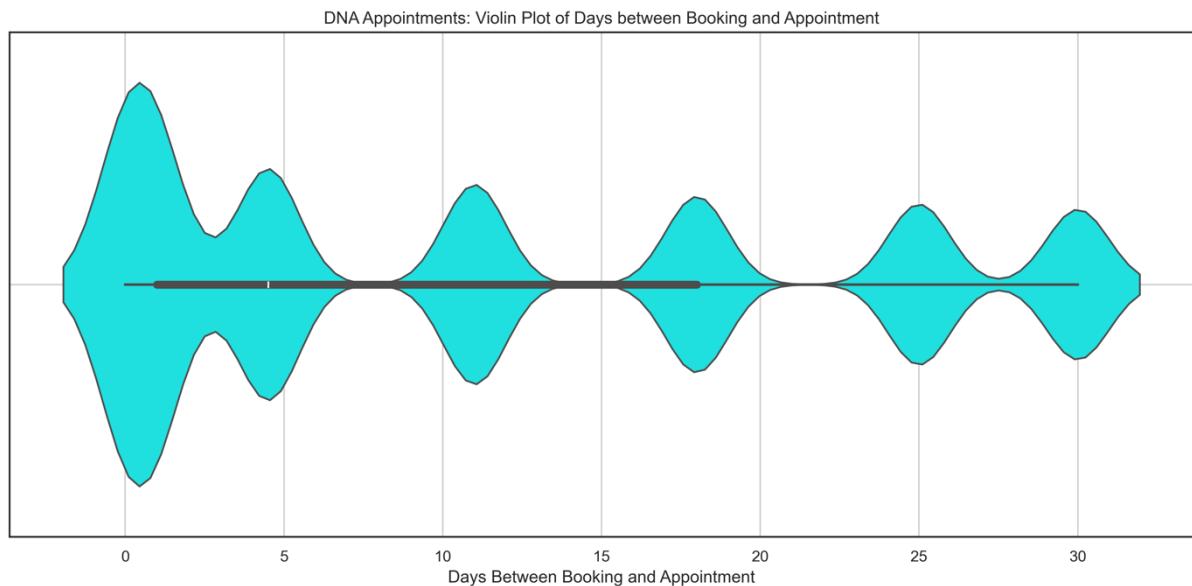
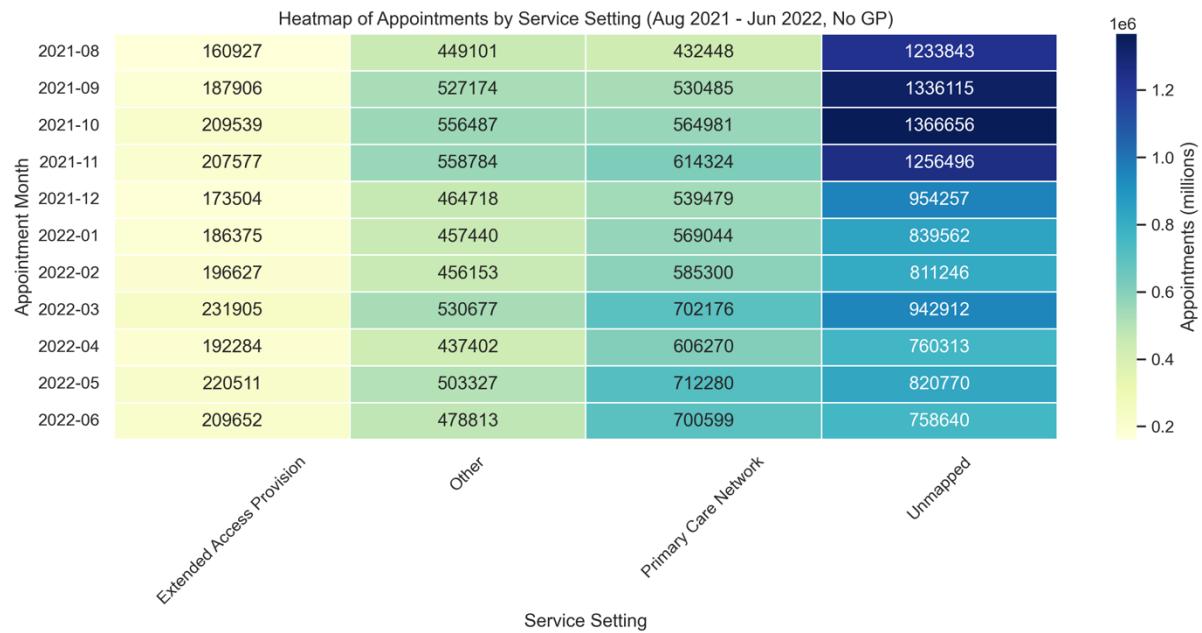


<https://PMC.ncbi.nlm.nih.gov/articles/PMC10884712/>

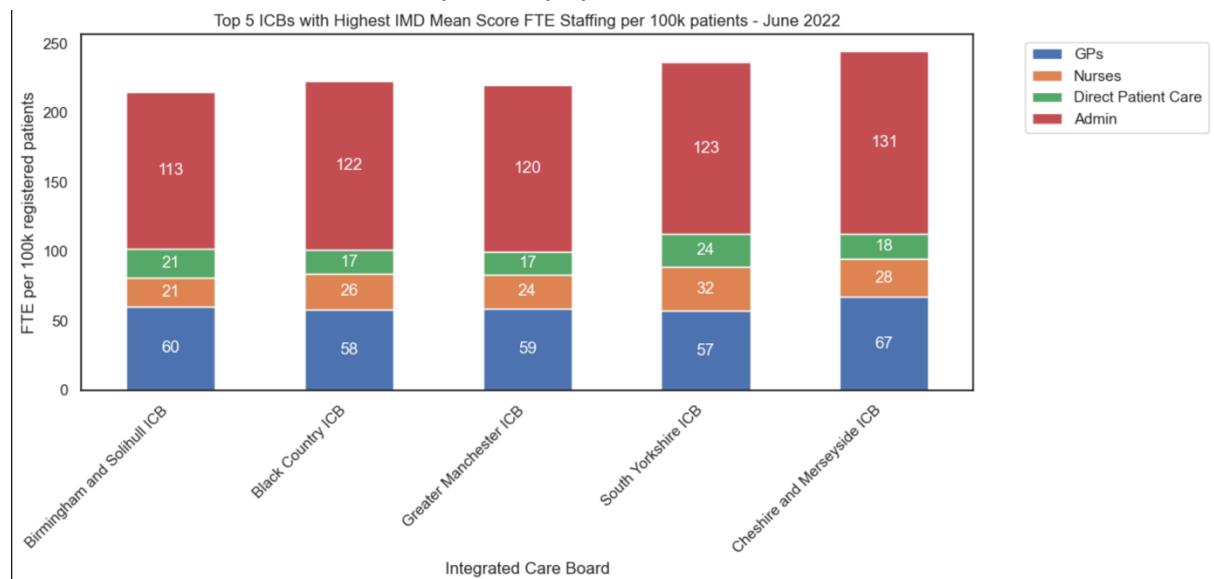
4. DNA appointments - There is a relationship (positive correlation) between Mean IMD Score (Index of Multiple Deprivation) and Total (Did not Attend) DNA Appointments - as the deprivation score increases, the number of missed appointments tends to increase as well.



## 5. Access is low. 1 Day wait time for appointments most likely to be attended.



6. Staffing composition for high IMD mean score ICBs do not address the complex and multi-faceted needs of the patient population.



<https://www.kingsfund.org.uk/insight-and-analysis/data-and-charts/health-inequalities-nutshell>

## Recommendations

## **Area: Appointment capacity allocation and utilisation**

## Recommendations:

- To deal with missed appointments, improve access by increasing EA provision capacity overall. This will enable more appointments in the optimal 1 Day timeframe.
  - EA provision can be telephone appointments where appropriate.
  - Other than the EA clinics soon to be delivered and funded as part of the Network DES, PCN may use streams of funding such as the Investment and Impact Fund (IIF) and the Capacity and Access Fund to improve access and adopt new technology.
  - Direct capacity to those who need it the most – ICBs with the highest IMD mean scores.
  - Increase capacity planning by 0.5% (in line with growth rate highlighted) every 6 months but in a targeted and efficient manner! Can adjust for seasonal variations (winter).

## **Area: Staffing capacity**

## Recommendations:

Increase Direct Patient Care for ICB areas with high deprivation levels to reflect the complex healthcare and social care needs of the patient population. Primary Care Networks may utilize their Additional Roles Reimbursement Scheme (ARRS) allocation to employ workforce with diverse skillset. Shift capacity away from GPs.

## **Area: Merits of external data source/leveraging data**

Recommendation:

- Leverage data/technology to improve patient experience and health outcomes. Reformulate marketing to enable hashtag tracking and feedback from patients – this will aid capacity planning and with tracking utilisation numbers.

## **Further actions**

- Sub-ICBs offer more localised healthcare offerings to their patient population as deprived areas tend to have high deviations in IMD score e.g. Brent in NHS North West London ICB. Perform analysis at this higher level of granularity.
- Post-June 2022, it is now easier to map ICB ONS codes to GP Workforce and List size data. Revisit analysis in 6 months' time to review effectiveness of recommendations, if these are implemented.

# Appendices

## Appendix A

### Key terms

- ICS, ICP, ICB: <https://www.england.nhs.uk/integratedcare/what-is-integrated-care/#:~:text=There%20is%20one%20ICB%20in,the%20ICP's%20integrated%20care%20strategy>
- Primary Care Networks (PCN): <https://www.england.nhs.uk/long-read/primary-care-networks-pcns/>
- Extended Access (EA) Provision: <https://www.england.nhs.uk/gp/investment/gp-contract/network-contract-directed-enhanced-service-des/enhanced-access-faqs/>
- Additional Roles Reimbursement Scheme (ARRS), Investment and Impact Funding (IIF), Capacity & Access Funding : See Network Directed Enhanced Service (DES) Specifications: [https://www.england.nhs.uk/wp-content/uploads/2022/03/B1963\\_i\\_Network-Contract-DES-Specification\\_171022.pdf](https://www.england.nhs.uk/wp-content/uploads/2022/03/B1963_i_Network-Contract-DES-Specification_171022.pdf)

## Appendix B

### Metrics generated and/or utilised

- FTE (Full-Time Equivalent) = 37.5hrs per week workforce capacity.
- IMD Mean Score: Ministry of Housing, Communities & Local Government, English indices of deprivation 2019 [www.gov.uk/government/statistics/english-indices-of-deprivation-2019](https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019)
- Patient List and Staffing type per 100,000 patients

```
# Standardise values to 100k population.

icb_capacity_merged['june22_appointments_100K'] = (icb_capacity_merged['total_appointments_june22'] / \
    icb_capacity_merged['icb_registered_patients_01jul22']) \
    * 100000).astype(int)

icb_capacity_merged['gps_june22_100K'] = (icb_capacity_merged['gps_fte_jun22'] / \
    icb_capacity_merged['icb_registered_patients_01jul22']) \
    * 100000).astype(int)

icb_capacity_merged['nurses_june22_100K'] = (icb_capacity_merged['nurses_fte_jun22'] / \
    icb_capacity_merged['icb_registered_patients_01jul22']) \
    * 100000).astype(int)

icb_capacity_merged['direct_patient_care_june22_100K'] = (icb_capacity_merged['direct_patient_care_fte_jun22'] / \
    icb_capacity_merged['icb_registered_patients_01jul22']) \
    * 100000).astype(int)

icb_capacity_merged['admin_june22_100K'] = (icb_capacity_merged['admin_fte_jun22'] / \
    icb_capacity_merged['icb_registered_patients_01jul22']) \
    * 100000).astype(int)
```

- Patient List Size growth rate Jan 2020 – July 2022

- Growth over 31-month period = 2.3954%
- Monthly growth rate:  $0.0765\% \text{ (} r = (1 + 0.023954)^{(1/31)} - 1 \text{)}$
- 6-month period growth rate: 0.461% ( $r = (1 + 0.023954)^{(6/31)} - 1$ )

## Appendix C

### List of key analysis performed

Sentiment analysis using Twitter data.

Seasonal decomposition of registered patient list size (Jan 2020 – June 2022). Note: A decision was made against generating a forecast of the patient list size using ARIMA as the patient list size exhibits seasonality. Steps to first transform the data to make it stationary would have to be undertaken and this was deemed beyond the scope of this analysis.

Investigating the link between deprivation levels and capacity allocation & utilisation June 2022.

Staffing capacity allocation and composition June 2022

EA provision analysis June 2022