

## **Assignment : Diagnostic Analysis using Python**

### **Introduction**

The National Health Service (NHS) was established in 1948 and delivers millions of appointments per year. Unfortunately, inefficiency costs are high due to missed appointments. This report aims to identify patterns and trends within NHS data to uncover why appointments are missed. This can lead to further action for the institution to adapt certain regions, locations, or appointment types to improve attendance rates. The report aims to solve this business problem:

- What are the underlying trends and patterns for the NHS regarding capacity and missed appointments?

### **Library Import**

- The zip file downloaded was saved onto a local disk. The import, clean, and analysis was performed in Jupyter Notebook. PEP 8 guidelines ensured code was legibly organised. Libraries imported were:

Python Library	Justification
Pandas	Powerful data manipulation and analysis library. Enables data structures to perform filtering, grouping, merging, and aggregation.
Numpy	Mathematical calculations and array operations.
Matplotlib	Data visualisation allowing high quality visualisations.
Seaborn	Used with Matplotlib as a supporting visualisation tool.
Datetime	Enabled conversion from object data types to parse data formats.

### **Import to Jupyter Notebook**

Data was imported using 'pd.read' function. Files that were csv type used pd.read\_csv, whereas excel workbooks were pd.read\_excel. An additional file (nhs\_names.csv – renamed 'rc') was created from ONS (<https://geoportal.statistics.gov.uk/>) allowing regional names. Files were assessed in shape and format using 'head()' and 'dtypes()' to check rows size and column titles.

### **File cleaning**

Copies of original files ensured any error with file manipulation could be backtracked. This was done using the 'copy()' method. Cleaning involved searching for null values, date formats, merging DataFrames and assessing categorical data for outliers:

Cleaning	Actual_duration.csv	Appointments_regional.csv	National_Categories.csv
Null values	None found using function 'isnull().any()'.	None found.	None found.
Parsed Data	Function variable 'pd.to_datetime' performed on column 'appointment_date'.	Column 'appointment_month' parsed to date format.	Column 'appointment_month' parsed to date format.
Merged DataFrames	'rc' was joined to the file using a left merge, only region code and name was added to actual_duration.csv.	Same as actual_duration.csv merge.	Same as actual_duration.csv merge.

<b>Categorical Outliers</b>	Unknown / Data Quality identified. This was not removed because the row still contributed to count of appointments.	Outliers found such as 'Unknown, and Data quality', unremoved with same reasons as actual_duration.	Outliers found such as 'Unknown, and Unmapped', unremoved with same reasons as actual_duration.
-----------------------------	---	---	---

### Insights and visualisation justification

#### Visualisation Rationale

- By Geography, Day of the week and Measure, simple bar charts were chosen as this is the most effective visualisation for comparison among small number of categories. When categories were numerous, a horizontal bar chart was selected for clarity.
  - Despite this, bar charts can struggle with too many categorical variables as bars are bunched. Furthermore, y axis differences can be misinterpreted when showing differences if not scaled correctly.
- By time, Pointplots show comparison over time. A rolling average removed seasonal patterns to uncover trends. By setting, the legend differentiated trends with colour. A dotted grid axis assisted y axis visual alignment.
  - A limitation is points will connect in straight lines, when there maybe micro peaks and troughs between points.
- For Attendance vs Wait time, a boxplot highlighted spread of distribution per wait time. This showed outlier values outside whiskers.
  - Boxplots can be misinterpreted for those that are unfamiliar with interpreting whiskers and the median.
- For Attendance vs Appointment setting, a scatterplot visualised differences per category. A legend showed different regions by colour. The palette was set to colourblind to support visually impaired.
  - This graphs limitation is points are condensed close together making it hard to identify differences per appointment setting. Further visuals should split out each setting to reduce the number of points shown.

### Insights and Recommendations

#### Geography (Section 3.1 in Notebook)

- The busiest region was the Midlands (18% of total appointments).
- The busiest ICB location was Northeast and North Cumbria (6%).
- The busiest Sub location was Northwest London ICB – W2U3Z (4%).
- Using Appendix 1, the ICB Northeast and North Cumbria is covering a huge area proportionally to other ICBs.
- Recommendation to investigate whether this ICB needs restructuring to spread resources more efficiently.

#### Day of the Week (3.2)

- The busiest day per week is Tuesday (21%).
- The beginning of the week is busiest, and slowly declines to Friday. Weekends have very few appointments.
- Recommendation to open hours more on weekends to alleviate weekday capacity.

#### Measure (3.3)

By Total Appointments:

- General practice makes up 91%.
- Care Related Encounters make up 87%.
- By National Category, the highest is General Consultation Routine with 33%.
- By Profession, 51% of appointments were GPs and 46% was other practice staff.
- Between Jan 20- June 22, there was a 91% attendance rate.
- Majority of appointments were Face-to-Face (59%) and Telephone (36%).
- The above demonstrate majority of appointments are face to face visits in general practices, with a large proportion being regular check-ups.
- Recommendations should investigate whether more face-to-face routine check-ups can be converted to e-consultations via phone or video call.

#### Time (3.4):

- Total Appointments are trending upwards in quantity.
- Peak seasonal appointments occur during winter and in particular December.
- Both summits in demand coincide with peak COVID infection peak periods so must be taken with caution as an outlier to general trends.
- Telephone consultations increased heavily during lockdown; however, they have slowly declined. This suggested as more social restrictions were lifted, patients were trending back to face-to-face visits than phone calls.
- Further investigation using qualitative data should question why telephone calls reduced, and whether more resource could be funnelled to virtual and telephone consultations to reduce face-to-face capacity.

#### By Comparison

- Between attendance and wait times, same day consultations had a 96% attendance rate, whereas more than 28 days had 73%. This insight suggests worse wait times cause more inefficiency with appointments.
- For appointments with wait times over 28 days, other practice staff contribute the most appointments. The NHS should investigate whether particular focus should be spent on this group to increase capacity to reduce wait times and therefore improve attendance.
- Lastly by appointment mode, telephone calls have the highest attendance rate suggesting an area that could alleviate face-to-face appointments.

#### Data Quality Recommendations:

- The data sets identified clear variations per ICB regarding data collection. For example, health care professions being classified as other practice staff. Therefore, insights must be taken with slight caution regarding the recommendations.
- A strong recommendation for the institution is to investigate ways to improve data collection and help define certain measures for ICB to utilise correctly.
- Lastly, qualitative methods should be used to enhance insights. This would assist understanding why ICBs record data differently.

Northeast and North Cumbria ICB

## Appendix 1 – Geographical Map of ICB locations

Website - <https://www.gbmaps.com/nhs-icb-maps/free-icb-maps.php>



### Integrated Care Board Boundaries (November 2022)

Integrated Care Board	Population
Bath and North East Somerset, Swindon and Wiltshire Integrated Care Board	923308
Bedfordshire, Luton and Milton Keynes Integrated Care Board	682311
Birmingham and Solihull Integrated Care Board	1358012
Black Country Integrated Care Board	1202528
Bristol, North Somerset and South Gloucestershire Integrated Care Board	969256
Buckinghamshire, Oxfordshire and Berkshire West Integrated Care Board	1016705
Cambridgeshire and Peterborough Integrated Care Board	859830
Cheshire and Merseyside Integrated Care Board	2228003
East of England Integrated Care Board	573299
Coventry and Warwickshire Integrated Care Board	897721
Derby and Derbyshire Integrated Care Board	1626492
Devon Integrated Care Board	1160998
Dorset Integrated Care Board	776780
Frimley Integrated Care Board	728827
Gloucestershire Integrated Care Board	640650
Greater Manchester Integrated Care Board	2937509
Hampshire and Isle of Wight Integrated Care Board	1540525
Herefordshire and Worcestershire Integrated Care Board	660601
Herefordshire and West Essex Integrated Care Board	646372
Humber and North Yorkshire Integrated Care Board	1804880
Kent and Medway Integrated Care Board	1763845
Lancashire and South Cumbria Integrated Care Board	3181028
Leicester, Leicestershire and Rutland Integrated Care Board	814372
Lincolnshire Integrated Care Board	670147
Mid and South Essex Integrated Care Board	963564
Norfolk and Waveney Integrated Care Board	1151283
North Central London Integrated Care Board	1247066
North East and North Cumbria Integrated Care Board	3458085
North East London Integrated Care Board	2036470
North West London Integrated Care Board	2235362
Northamptonshire Integrated Care Board	915646
Nottingham and Nottinghamshire Integrated Care Board	1511928
Shropshire, Telford and Wrekin Integrated Care Board	506737
Somerset Integrated Care Board	563851
South East London Integrated Care Board	1818226
South West London Integrated Care Board	923837
South Yorkshire Integrated Care Board	1774079
Staffordshire and Stoke-on-Trent Integrated Care Board	1139794
Suffolk and North Essex Integrated Care Board	855426
Surrey Heartlands Integrated Care Board	1367876
Sussex Integrated Care Board	1717454
West Yorkshire Integrated Care Board	2406920

### For Designers

Buy a bundle of vector editable files showing all the NHS Integrated Care Board boundaries along with the Local Authority Boundaries.

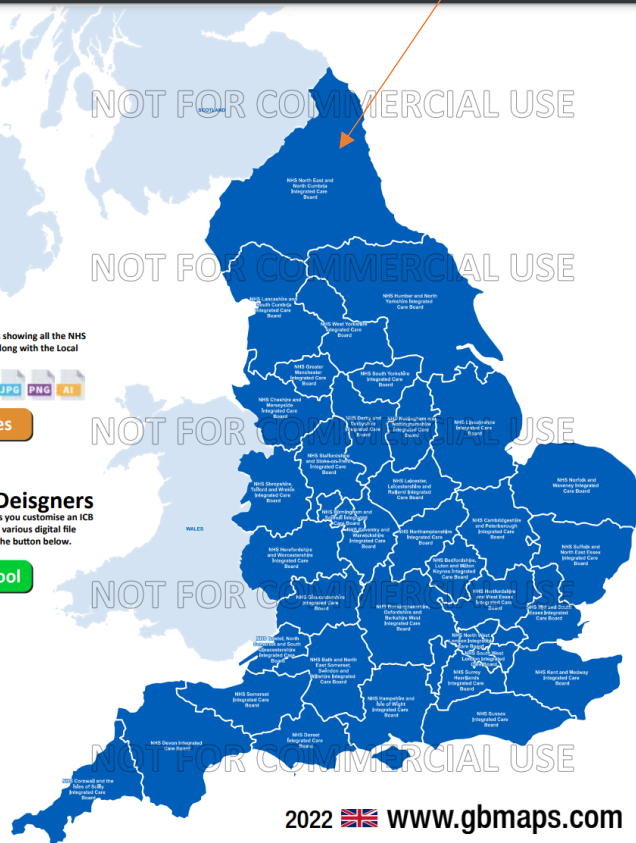


Buy Editable Files

### For Wannabe Designers

We've built a Mapping Tool that lets you customise an ICB map online and then download it in various digital file formats. Give it a try for free using the button below.

Map Colouring Tool



2022 [www.gbmaps.com](http://www.gbmaps.com)

NorthEast and North Cumbria ICB :



2023 - [www.gbmaps.com](http://www.gbmaps.com)

Download this map as a FREE pdf file: [FREE NHS North East and North Cumbria Integrated Care Board PDF Map](#)

