

# NHS

## Data Analysis Overview

# Background

1

During my analysis of the National Health Service (NHS) datasets, I aimed to help the NHS to improve its appointment attendance whilst creating an efficient and effective healthcare service.

To reach these objectives I analysed:

- How are the NHS services functioning at the moment?
- What trends are there across the dataset? How do NHS services differ across different months and locations? What might be the reason for this?
- Which areas have poor data quality?
- What is the utilisation of services? Will this change in the future?
- How is the public discussing the NHS online?

## Analytical Approach

Here is the analytical approach I followed to gain insights.

### Importing

I imported different libraries such as seaborn, pandas, numpy, matplotlib, and Holt's method for exponential smoothing into the notebook. I set all figures in the notebook to a desired size and style and set the desired resolution for these figures.

## Sense Checking

Theses are methods used to gain an understanding of the structure of the different data sets

- `.columns`
- `.shape`
- `.dtypes`
- `df[column_name].sum`

## Data Cleaning

- `.isna().any(axis=1)`: This method clearly shows whether or not there are missing values in a whole data set.
- `pd.to_datetime`: to transform the dates into date time format
- I used the `.unique` function to see if the month or date columns in the data set had any outliers.
- I decided not to try to remove outliers from the 'count of appointments' because the metadata states the data is not definitely an accurate depiction of what the data looks like in reality. Therefore, removing outliers wouldn't necessarily make the data more accurate.

## The Most Recorded Values

- To see the top recorded locations in a column I used the `.values_counts` function to see the number of instances a name appears in a row.
- The locations that came up the most came up the same number of times. `[:10]` helped me to distinguish what the top locations were.

## Creating Subsets

- I use the `groupby()` function to group certain columns with the `.agg("sum")` of the 'count of appointments'. This creates a new data frame that shows the count of appointments in a certain chosen category. The `.rest_index()` function rearranged the x-axis in chronological order during this process.
- I used the `loc.` function to choose a column and then the `isin.` function to choose a value within that column. This helped to create a new subsets. The NOT bitwise operator (`~`) combined with `str.contains` can be used to remove a value in a column from a data set.
- I used the `.merge()` function to combine data frames depending on common values in a certain column.

## Iterating Across Data Frames

I iterated across a data frame to calculate the percentage of a certain value appearing in a column in a certain month or date. I did this firstly by grouping the appointment month or data into a new data set. I then created an empty data frame to store the total number a certain value appears in a column as well as the total number that all values appear in a column.

To find these values I iterated over the month or date data frame. I then looked at each entry for that month and used a list comprehension to count all appointments that were marked as a certain value in a column. I also did this to find the number of total values in the column for a certain month or date. I then placed this information into a new data frame. This helped me to see, for example, how the proportion of GP appointments changed over time compared to other service settings.

## Forecasting

To create a forecast, the rolling mean is used to track how the average count of appointments for a certain category changes over time. This is done using the `mean. function` with a specified rolling widow. Then Holt's method for exponential smoothing uses an exponentially weighted moving average to predict future values.

## Searching for trending hashtags

Here I use a list comprehension to create an array of arrays, where each array in the array is a tweet that is split by the space character (i.e. each element in the array is a word from that tweet). I then iterate over each word and check whether there is a '#' character in that word. If there is, then that is a tag and can be appended to my tags array, otherwise, we continue with the loop.

## Searching for Specific Words in Tweets

Creating a user defined function in combination with the `apply()` function makes it possible to search for certain words in the Twitter data set.

## Visualisations

I used bar charts to compare the size of different categories in a data set. I used the `palette='deep'` to help distinguish between different values that were not chronologically connected.

Occasionally I used grouped bar and line charts. The `twinx()` function is used to combine charts into one with the same x-axis but different y-axis. This makes it possible to find similarities between two different pieces of data.

I then used line charts to demonstrate how the sum of the count of appointments in different categories changed over time.

Scatter plots were then used to demonstrate how different values within a category were more or less popular in certain locations. This is the easiest way to identify trends across different categories in different locations. They were also used to identify trends across many different data points.

To visualise the actual duration of appointments and the time period in between appointments in graphs I used a blended colour palette with a gradient. This is because the time values were chronologically ordered. This colour scheme made it easier to identify different trends on the charts.

## Key Patterns and Predictions

- Primary Care Networks are gaining traction.
- The percentage of service settings that are GP appointments increases in the winter. More people demand GP appointments during this time.
- There is a high demand for planned clinics during the autumn.
- It takes the longest to book an appointment in NHS North East and North Cumbria, NHS West Yorkshire. Perhaps there is a need for more staff in these areas?
- Telephone appointments are becoming less popular whereas video appointments are becoming more popular. Invest in improving video services.

- Prepare to increase staff capacity by mid 2024.
- Attendance rate of appointments worsens during the winter months.
- Further investigate why areas such as Greater Manchester and South East London have the lowest attendance rate.
- The rise of AI is creating new opportunities in healthcare which could help the NHS to reach its objectives.



Helping you to make  
commitment to quality  
of care.