

I am part of a team of data analysts that was approached by the National Health services (NHS). NHS have a problem where they are incurring significant costs, when patients miss general practitioner (GP) appointments. Penalising patients for missing appointments can affect the poorest and most vulnerable in the community. NHS wants a data driven model to understand the reasons for missed appointments, and how it can be reduced to save money, which would benefit financially and socially.

The two main questions posed by the NHS, that I will be investigating are:

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

### **Exploratory Data Analysis(EDA)**

If we want to explain EDA in simple terms, it means trying to understand the given data much better, so that we can make some sense of it, to enable us to answer some key questions. The more we explore the data, the more the insights we draw from it. As a data analyst, almost 80% of our time will be spent understanding data and solving various business problems through EDA.

### **Data Sourcing**

Data Sourcing is the process of finding and loading the data into our system. Broadly there are two ways in which we can find data- Private Data and public data.

In this project we are using public data that was given to us to explore: 'appointments\_regional', 'actual\_duration', 'tweets' were all CSV files, while 'national\_categories' was an excel file.

The data was imported - "import pandas as pd, import matplotlib.pyplot as plt, import seaborn as sns". Then we created data frame out of the CSV files – by coding pd.read\_csv. For the excel file it was pd.read\_excel .

We then used Describe to find out key statistics like mean, count etc.

Pandas **dataframe.info()** function is used to get a concise summary of the dataframe.

Head function was used to find top 5 rows, and tail function for the bottom 5 rows. The shape method was used to analyse the data and its variants, and to realise the volume of the data.

We looked at all datasets to make it look better, if need be.

We run metadata\_nhs file to see what each columns meant. We saw actual duration in text format, we expected it in numeric format. We removed nulls.

We use pandas as library. Pandas great for data wrangling and transformation. We used read csv function to load csv files and read.excel to load excel files.

We have used matplotlib and seaborn to visualise and make the plot.

The graph wasn't aligned properly and we used plt.subplot function sns.set function. "fig, ax = plt.subplots()

```
sns.set(rc={'figure.figsize':(14, 8)})
```

```
sns.set_style('whitegrid')
```

We used seaborn library and matplotlib library to make it more aesthetically appealing.

unique() function was also used to **find the unique elements of an array and returns these unique elements as a sorted array.**

### Visualisation and insights:

Why did we use line graph? It is Easier to see. If you are trying to visualise a certain metric over time- a line graph is the best utilisation, as its simple to understand and you can add colour.

On my notebook I displayed three line graphs. First one is showing 'Appointments per month per service setting with number of appointments'. Second one is showing 'Appointments per month per context type with number of appointments'. Third line graph is showing 'Appointments per month per national category with number of appointments'. All three graphs show that appointments rise steadily from August to November. Then in November and December it falls rapidly.

We would expect appointments to be high in November and December due to Christmas/New year parties, where people tend to be drunk and hence more risk to injury etc. In January it goes up rapidly so that's something the NHS can investigate why.

Again appointments go up in January and there is a big jump from February to March. After March it decreases. Start of November and start of March are the highest appointments, but then they both fall massively by the end of the month. Again why is that? Something the NHS can investigate further.

First graph shows that most appointments are in general practice. The rest are similar to each other. Second graph shows care related encounter has the highest appointments, and rest are similar. Third graph shows General consultation acute has the highest appointments, and rest are similar.

We then investigated what are the top trending hash tags (#) on Twitter related to healthcare in the UK?

I used seaborn to plot a bar chart for hash tags. The top trending hash tags were #healthcare , #Healthcare, #health, #Healthcare. Important to note that these can be all ranked 1<sup>st</sup> under one topic of healthcare. #AI is ranked 2<sup>nd</sup> and it seems people are talking

more about advent of AI into healthcare, and what role it will play. Job and strategy is ranked 4<sup>th</sup> and 5<sup>th</sup> which shows people are keen to know what jobs are available within healthcare. Digital health is ranked lowest but it shows people are looking at ways to enhance the efficiency of healthcare delivery, and to make medicine more personalized, direct and precise.

### **Patterns and predictions:**

There were less patients fighting for GP slot in NHS Staffordshire and Stoke-on-Trent ICB.

NHS North West London ICB had the highest patients fighting for GP slot. NHS Staffordshire and Stoke-on-Trent are doing something right, therefore NHS needs to investigate further and mimic what they are doing across all areas.