Diagnostic analysis using Python

NHS is a public healthcare system of England. The NHS incurs significant, potentially avoidable, costs when patients miss general practitioner (GP) appointments. Therefore, reducing or eliminating missed appointments would be beneficial financially as well as socially. The government needs a data-informed approach to deciding how best to handle this problem. At this stage of the project the two main questions posed by the NHS are:

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

Data Analysis was done on three datasets actual_duration.csv, national_categories.xlsx, appointment_regional.csv. Actual duration csv contained 137793 rows and 8 columns. No NaN values were found in any of the columns. As per metadata, duration less than 1 minute or greater than 60 mins are grouped as Unknown / Data Quality, this amount to 20161 rows, which is 15% of the appointments. This data if treated as outlier may give some insights of where duration is taken more. Appointments Regional data did have any null values, it contained 596821 rows and 7 columns.
From this dataset we found Unknown values in below columns
Unknown Appointment Status: 201324 => 33 % of records
Unknown HCP Type: 201324 => 33 % of records
Unknown Appointment Mode: 79147 => 13 % of records.

National Category data set has 817394 rows, 8 columns. No Unknown or null values found.

The below results were arrived, datasets were created from csv & xls using
DataFrame read_csv & read_excel functions.
Null check is done using is nc[nc.isna().any(axis=1)]
metadata of the datasets are found using describe() & info()
counts were arrived using value_counts() and by filtering

Number of location : 106
High number of records (Top 5):
NHS North West London ICB - W2U3Z,
NHS North East London ICB - A3A8R,
NHS Kent and Medway ICB - 91Q,
NHS Hampshire and Isle Of Wight ICB - D9Y0V,
NHS South East London ICB - 72Q

Number of Service Settings: 5
Number of Context Types: 3
Nu Number mber of National Categories: 18
of Appointment Status: 3 (Attended, Unknown, DNA)

National Categories were analyzed, visualizations were created using seaborn.
Line plots were created for number of appointments per month for service settings, context types, and national categories.This was done by grouping the data set by appointment_month & required.The data insights from the graph shows the no of appointments increase by season, General Practice forms the major service used.
Care Relate Encounter forms major chunk of context type, General Consultation forms major national category.

Seasonal data are collected by grouping by required category & appointment_month and count_of_appointments were summed to get for specific category.
It's clear from the data that General Practice has more appointments than any category irrespective of the season.

Tweets related data were analyzed using dataframe methods, describe(), dtypes.
twitter full text was filtered from dataset, hashtags were found from the full text using string operations. This data was converted to dataframe and no of occurrence of each tag were found.Hastags & count were put into a dataframe and bar chart was created to find which hashtag was trending. Bar chart for least used hashtags were also plotted but these hashtags were filtered so that it's related to medicine. From the tweets & hashtags we can find the health care & insurance forms the key hashtags used while tweeting. People are more interested to tweet about these tags, moreover the text on these tags can provide insights on what patients are looking for & which service setting.

Patterns were found based on visualization, here is considerable no of appointment is in DNA or Unknown status. This needs to be looked into, there is a possibility if every GP follows common workflow the Unkown status may come down. End & beginning of the year appointments are decreasing, otherwise increasing trend, except Unknown (this could be due to workflows followed by practice). There has been reasonable increase in telephonic appointments. The box plot was right skewed, with lots of outliers. When removed the outliers using showfliers=False option, the plots looked better but still skewed towards right. The mean looks positive in summer but more negative after the summer. When removing the GP, out of service setting it looks much better & positive after removing outliers. This provides insight that major issue is in General Practice; this is the place where the appointments go beyond a reasonable time causing more issues to NHS.

In conclusion, based on the visualizations, focus of workflows can improve more meaningful data, can give better insights. The no of appointments is seen more in big cities / counties., where the utilization is really high. The increase in appointments is majorly seen in Autumn / Winter. There has been reasonable increase in telephonic appointments which could help to resolve appointments quickly. Workflow of General Practice could be improved as this area generates more appointment than any other service.

.