Diagnostic Analysis using Python

Business scenario

The NHS is facing challenges in managing its growing patient demand and optimising its resource utilisation across a wide network of services. The increasing population in England requires careful budget allocation to expand NHS infrastructure while ensuring the existing resources are used efficiently. A particular focus is on understanding the causes of missed GP appointments, as this results in avoidable costs.

1. Provide actionable insights to inform NHS decisions.

2. Help NHS scale infrastructure for growing demand.

3. Guide NHS budget allocation through utilisation trends.

4. Analyse Twitter data to inform public sentiment and budget decisions.

Analytical approach:

Downloading the data

I imported the datasets into the Python environment using pandas. These datasets were provided in different formats, including CSV and Excel files. I used pd.read_csv() to load the CSV files and pd.read_excel() to load the Excel file. The files included the following:

actual_duration.csv: This dataset contains information about the actual duration of appointments. appointments_regional.csv: This dataset includes regional data about appointments.

national_categories.xlsx: An Excel file containing national category data, useful for aggregating regional data. tweets.csv: This dataset contains tweets, useful for twitter insights.

Data cleaning and quality check:

Checked for missing values and inconsistencies. Standardised 'appointment date' to datetime format. Merged datasets for a unified view and ensured consistent naming conventions. Detected and addressed outliers.

I selected visualisations based on the need to clearly communicate trends and insights. Time Series Analysis (Line Chart) for appointment trends over time, identifying seasonal patterns and demand peaks for capacity planning and resource allocation.

Bar Charts were used to compare appointments across service settings, regions, and categories, as well as to analyse Twitter data, helping identify trends and inform public sentiments.

Exploratory data analysis:

Actual duration

The large range (15,399) and moderate IQR (1,427) with a high median (696) indicate a fairly distributed appointment duration, with minimal impact from outliers. Appointments Regional

National categories

The wide range (16,589) and small IQR (121) show most data points are clustered, with the median (25) indicating many categories have low appointment numbers.

Observations about the data dispersion

All three datasets are heavily skewed, with standard deviations much larger than the means, indicating extreme values or outliers.

Outlier Visualisation:

Extreme values exist that reflect real trends, like high demand or COVID-19.

Observations about the data dispersion

Actual Duration (ad): Min: 1 Max: 15400 IQR: 25% = 194, 50% = 696, 75% = 1621

   2. Appointments Regional (ar):

Min: 1 Max: 211265 IQR: 25% = 7, 50% = 47, 75% = 308

   3. National Categories (nc):

Min:1 Max: 16590 IQR: 25% = 7, 50% = 25, 75% = 128

Outliers, like Covid-related events, have minimal impact on summary stats (mean, standard deviation) due to the large dataset, so I've ignored them.

2. How many locations are there in the data set?

 The data shows the geographical spread as 106 locations. This helps evaluate utilisation efficiency

3.Five locations with the highest number of appointments?

Identifying these areas helps optimise resource allocation for staffing and capacity

```
Top 5 locations with the highest number of appointments:
                                    Location  Number of Appointments
0            NHS North West London ICB - W2U3Z                12142390
1            NHS North East London ICB - A3A8R                 9588891
2                  NHS Kent and Medway ICB - 91Q               9286167
3  NHS Hampshire and Isle Of Wight ICB - D9Y0V                 8288102
4               NHS South East London ICB - 72Q                7850170
```

4.How many service settings, context types, national categories, and appointment statuses are there?

This guides resource allocation, improves efficiency, and highlights areas for better patient care.

```
'Number of service settings: 5'

'Number of context types: 3'

'Number of national categories: 18'

'Number of appointment statuses: 3'
```

5. Between what dates were appointments scheduled?

Time-based analysis reveals seasonal trends, aids resource allocation, forecasts demand, and improves scheduling.

```
Appointments in 'ar' were scheduled between: 2020-01-01 00:00:00 and 2022-06-01 00:00:00
Appointments in 'ad' were scheduled between: 2021-12-01 00:00:00 and 2022-06-30 00:00:00
Appointments in 'nc' were scheduled between: 2021-08-01 00:00:00 and 2022-06-30 00:00:00
```

6.Month with the highest appointments:

November 2021, indicating a peak, which indicates winter surge but also the Covid Pandemic

7. Aggregate appointments per month by service settings, context types, and national categories.

GPs have higher appointment volumes.

8. Analysis of twitter data

Category	Total Count of Conversations
Health-Related	984
AI-Related	45
Job-Related	38
Strategy-Related	31
Social media	25

health-related tweets are the highest.

Data Visualisation:

Visualisation	Insights	Recommendations
1 Visualisations for each season.	Fluctuating demand across service settings. -Seasonal trends showing peaks for some settings	**Resource Allocation**: Ensure adequate staffing during peak seasons. **Seasonal Planning**: Adjust schedules and promote off-peak settings. **Data Monitoring**: Track trends for improved forecasting.
2. Appointment count by status and time between booking and appointment.	Most appointments confirmed; fewer cancellations/no-shows. Shorter booking intervals indicate urgency. Seasonal peaks in demand.	**Improve Scheduling Efficiency**: Optimise appointment times. **Reduce Cancellations and No-shows:** Use reminders and flexible rescheduling. 3. **Address Seasonal Demand Peaks:** Adjust staff availability during surges.
3.Comparison of service settings	Demand patterns revealed across service settings. Seasonal peaks/valleys in appointments.	**Optimise Resource Allocation:** Balance resources between underused and overused settings. **Seasonal Staffing**: Adjust staffing to meet seasonal demand fluctuations.
4. Comparison of appointment status (GP, non-GP, grouped)	GP appointments dominate with higher seasonal demand. Non-GP services show stable, lower numbers. "Unknown" services have data gaps.	**Allocate Resources for GP Demand:** Ensure GP services are adequately staffed. **Address Data Gaps:** Improve data collection for "unknown" services. **Promote non-GP Services:** Balance workloads by encouraging use of non-GP services.

5.Twitter data analysis for future investment	Assessing relevance, quality, and insights of Twitter data. Helps improve service delivery and understand customer sentiments and improve communications with the public.	**Invest in Twitter Data:** Twitter provides immediate, unfiltered public sentiment, offering insights into what people think about NHS services, policies, or health trends as they emerge.

Actionable recommendations to NHS

- Plan for seasonal staffing peaks and flexible workforce models to manage high GP demand.
- Launch campaigns to raise awareness of primary care, extended access, and unmapped services.
- Develop a Seasonal Appointment Management System and implement predictive scheduling for high-demand periods.
- Boost telephone/video consultations to reduce in-person appointment pressure.
- Use Twitter data to monitor public sentiment and emerging health trends for timely decisions.
- Introduce appointment reminders and analyse no-shows to improve attendance.
- Monitor and respond to negative feedback in real-time, track health trends.
- Increase GP availability and prioritise urgent appointments during seasonal spikes.
- Utilise other practice staff for non-urgent cases during high-demand periods.
- Implement data-driven scheduling and expand telehealth for non-urgent visits.
- Improve resource allocation, data quality, and forecasting for better service delivery.
- Run targeted campaigns, streamline urgent care, and focus on winter flu management.
- Respond to feedback and adjust services to improve public perception.
- Allocate more resources to NHS North East and North Cumbria; address access issues in NHS Shropshire.
- Train other primary care staff to handle traditional GP's tasks to ease the pressure on doctors.

References: integrated care board (ICB)(Lancashire and South Cumbria 2023) GP Practice New 2022.

NHS Digital 2022

Harvard University - CS50: Introduction to Computer Science

Publication Link: CS50x - Introduction to Computer Science Stack Overflow.

Appendix: **Screenshots and visualisations.**