

COURSE 2

Diagnostic Analysis using Python

ANAND JOSHI

22/05/2023

Word count (excluding “Contents”, “References”, “Appendix”): **1100**

Contents

CONTEXT	1
ANALYTICAL APPROACH	2
VISUALISATIONS & INSIGHTS	6
PATTERNS & PREDICTIONS	7
REFERENCES	9
APPENDIX	10

Context

Business Scenario

The National Health Services (NHS) is experiencing significant costs due to missed appointments. A team of data analysts has been contracted to better understand the reasons for missed appointments and develop a data-informed approach to reducing or eliminating them.

Business Problem

The NHS needs to identify the reasons for missed appointments with GPs in order to develop effective strategies for reducing or eliminating them. Specifically, the NHS wants to know whether the current staff and capacity in the networks are adequate, and what the actual utilisation of resources has been. By addressing these questions, the NHS hopes to improve patient outcomes and reduce costs associated with missed appointments. Additional questions will also be considered in the analysis such as:

- Are there any recurring patterns or trends in the timing of missed appointments?
- Among different healthcare profession, which profession has the highest number of missed appointments?

Analytical approach

GitHub repository

GitHub was used for this project due to it being easy for developers to share code files and collaborate with fellow developers on open-source projects. (See Jupyter Notebook for URL)

Python in Jupyter notebook

Following a structured approach to import, clean, and analyse the provided data sets. To accomplish this, I made several decisions regarding the choice of libraries, functions, and variables:

1. *Data Import* – Using the pandas library, a powerful data manipulation and analysis tool, to import the data sets, and Numpy for mathematical operations. Pandas provides efficient data structures and functions for handling structured data, making it an excellent choice for this task. I loaded the data sets into pandas DataFrames using the ‘pd.read_csv’ and ‘pd.read_excel’ functions. Then, sense checking the data using the columns(), shape(), dtypes() functions was performed before proceeding.
2. *Data Cleaning* – Cleaning the data to ensure its quality and integrity is important. This involved identifying and handling missing values, inconsistent formatting, and outliers. I used pandas' built-in functions such as ‘isnull()’ and ‘dropna()’ to identify and handle missing values appropriately.
3. *Data types* – Throughout the analysis, I made careful decisions regarding the selection of variables to focus on, considering factors such as their relevance to the problem statement. I also performed any necessary data transformations,

such as converting date strings to datetime objects using the 'datetime' library, to enable accurate analysis and visualisation.

4. *Exploratory data analysis (EDA)* – To gain insights and identify patterns, I used descriptive statistics, i.e. 'describe()' such as mean, median, and standard deviation, to summarize the data and understand its distribution.
5. *Visualisation* – Employing data visualisation libraries like Matplotlib and Seaborn to create meaningful line plots, bar plots, and boxplots to visualise trends and relationships within the data.
6. *Variables* – To specifically address this problem, I analysed relevant variables. This analysis helped in understanding the problem statement and suggested potential areas for intervention or improvement.

The decisions I made with data ingestion, wrangling, regarding the choice of libraries, functions, and variables were guided by the requirements of the given scenario and aimed at providing meaningful recommendations to tackle the problem statement for the NHS. (See Jupyter notebook provided for an in-depth pythonic code analysis of the above points).

Visualisation and insights

Visualisations

Rational:

- *Lineplot* – Used to visualise data that is continuous/sequential in nature. It allows us to observe changes, fluctuations, or correlations in the data over time or across different categories.
- *Barplot* – Ideal for visualising discrete data or data grouped into categories. The height or length of each bar represents the frequency or count of a particular category, allowing for easy comparisons between different categories. Barplots are effective for showing relationships, comparisons, or rankings among categories.
- *Boxplot* – They display important statistical measures such as the median, quartiles, and potential outliers. Boxplots are particularly useful for detecting skewness, variability, and the presence of outliers in the data.

Insights

There were many visualisations presented in the analysis of my Jupyter notebook. Relevant to the business objectives are:

1. Lineplot to show time between booking and appointments – Well-suited for illustrating the time between booking and appointments because it allows for the visualization of trends and patterns over time. By plotting this graph, it becomes easy to identify any variations or changes in the time gap between bookings and appointments. (See Fig.1 in Appendix 7)

2. Lineplot to show appointment by attendance status – Great choice to display appointments by attendance status as it can show the change in attendance over time. By plotting the graph, it becomes straightforward to observe any trends or patterns in attendance behaviour.
3. Lineplot to show monthly capacity utilisation – An effective way to demonstrate monthly capacity utilisation as it can display the changes in utilisation rates over time. By plotting this graph, it becomes easy to identify any seasonal variations or overall trends in the capacity utilisation throughout the year. (See Fig 1 in Appendix 3)
4. Barplot of Twitter data – It allows for easy comparison and identification of the most popular or trending hashtags per frequency.
5. Boxplot to show Spread of Service Setting – ideal for showcasing the spread of service settings because it provides a clear visualisation of the distribution, central tendency, and variability of the data. (See Fig. 1 & 2 in Appendix 5)

Patterns and predictions

General Practice emerges as the service setting with the highest number of appointments consistently. This consistent demand for General Practice services underscores the importance and popularity of primary care in addressing patients' healthcare needs. In terms of national categories, General Consultation Routine exhibits the most appointments. This suggests that routine consultations for general healthcare concerns are prevalent, highlighting the significance of primary care in addressing common health issues and providing comprehensive patient care. (See Fig. 1 in Appendix 2).

There was a slight decline in GP appointments between April 2020 to May 2020, likely due to the initial impact of the COVID-19 pandemic and related restrictions. However, appointment numbers steadily increased back to normal levels after that, indicating the resilience and adaptability of the healthcare system in responding to changing circumstances. (See Fig. 1 in Appendix 3). The NHS can accommodate a maximum of 1,200,000 appointments per day. There were some instances where the utilisation of resources came close to surpassing the threshold, but didn't. This was between October 2021 to January 2022. (See Fig. 1 in Appendix 3).

The time between booking and appointment varied throughout the observed period. It reached its highest point in December 2020, possibly reflecting holiday-related delays and increased demand. Conversely, it was at its lowest in March 2022, suggesting improved efficiency in appointment scheduling and reduced waiting times. (See Fig. 1 in Appendix 4)

Recommendations on areas for further exploration include the primary reasons/barriers for missed appointments by patients, common demographic factors or socio-economic variables

associated with missed appointments, and exploring the factors influencing service settings and availability.

References

- FourthRev, LSE_DA201_Data Analytics using Python_C1_2023, Available from: (fourthrev.com)
- Tutorial republic, Available from: [Python Tutorials – Real Python](#)
- Python Programming, Available from: [Python Programming Tutorials](#)

Appendix

1.

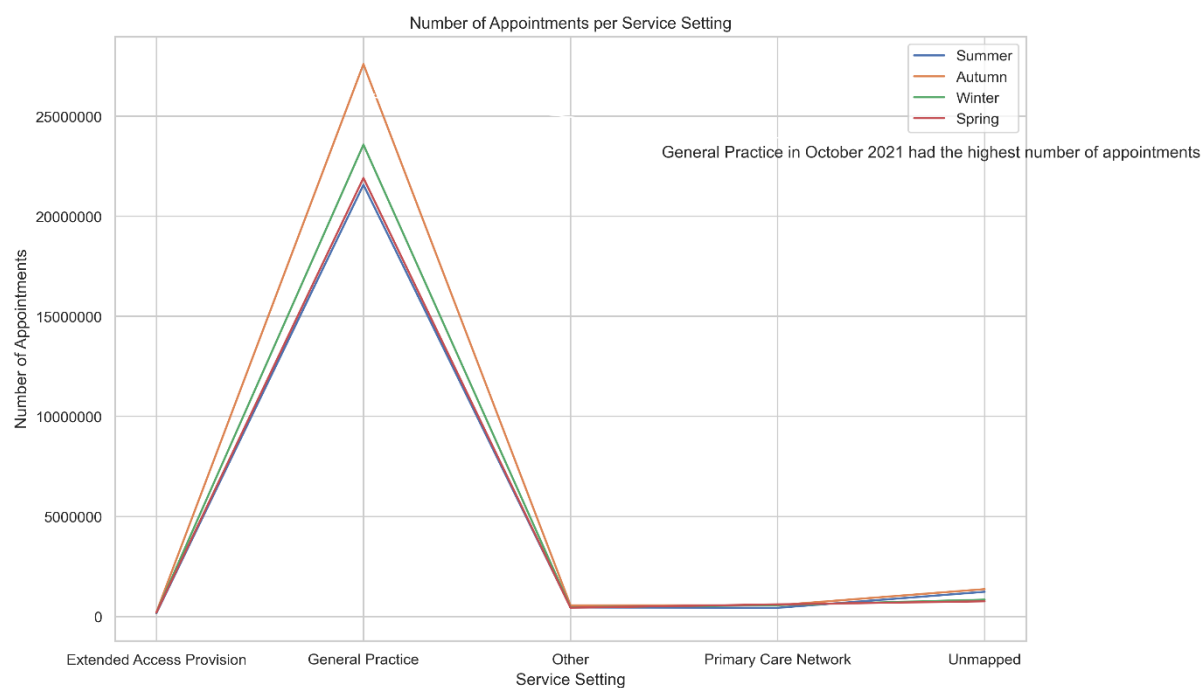


Fig. 1

Number of Appointments per Service Setting

Shows General practice in all seasons having the greatest number of appointments, with October 2021 at it's peak.

2.

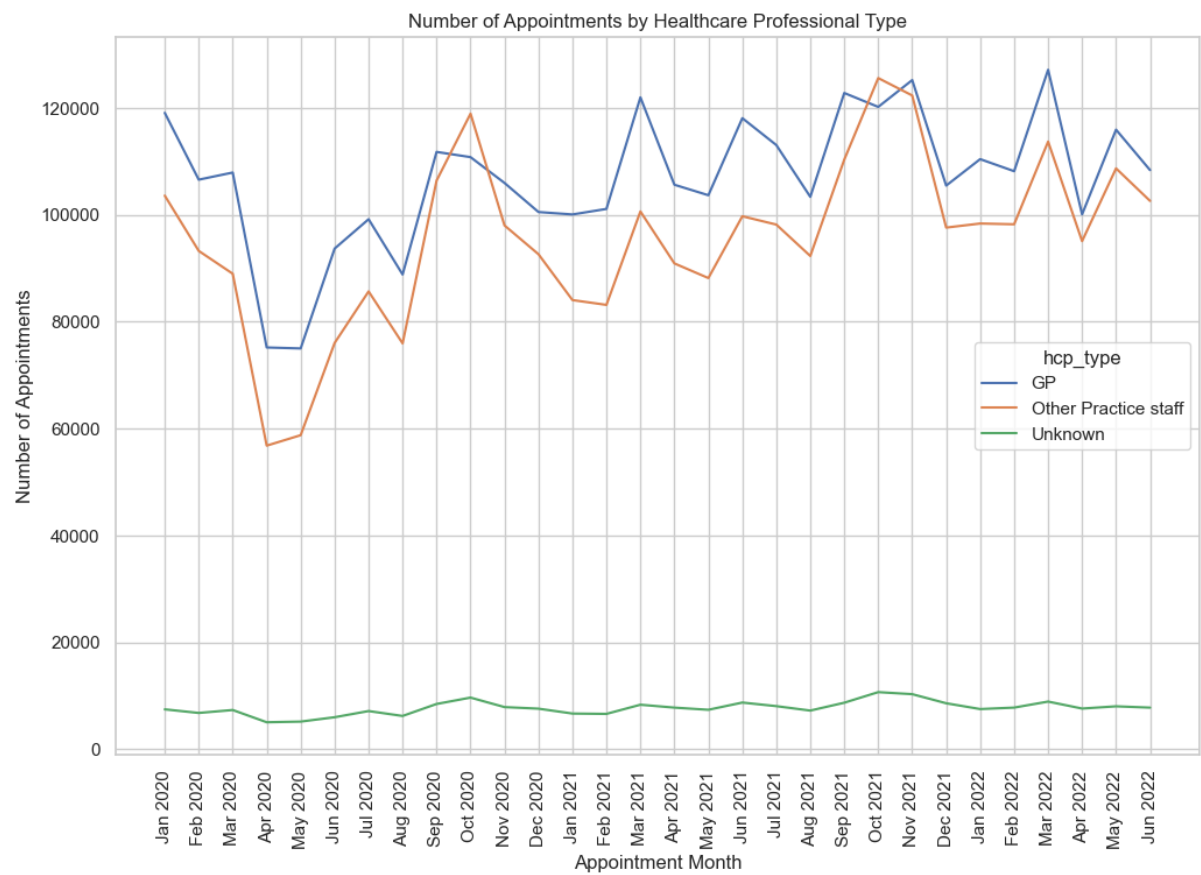


Fig. 1

Number of Appointments by Healthcare Professional Type

3.

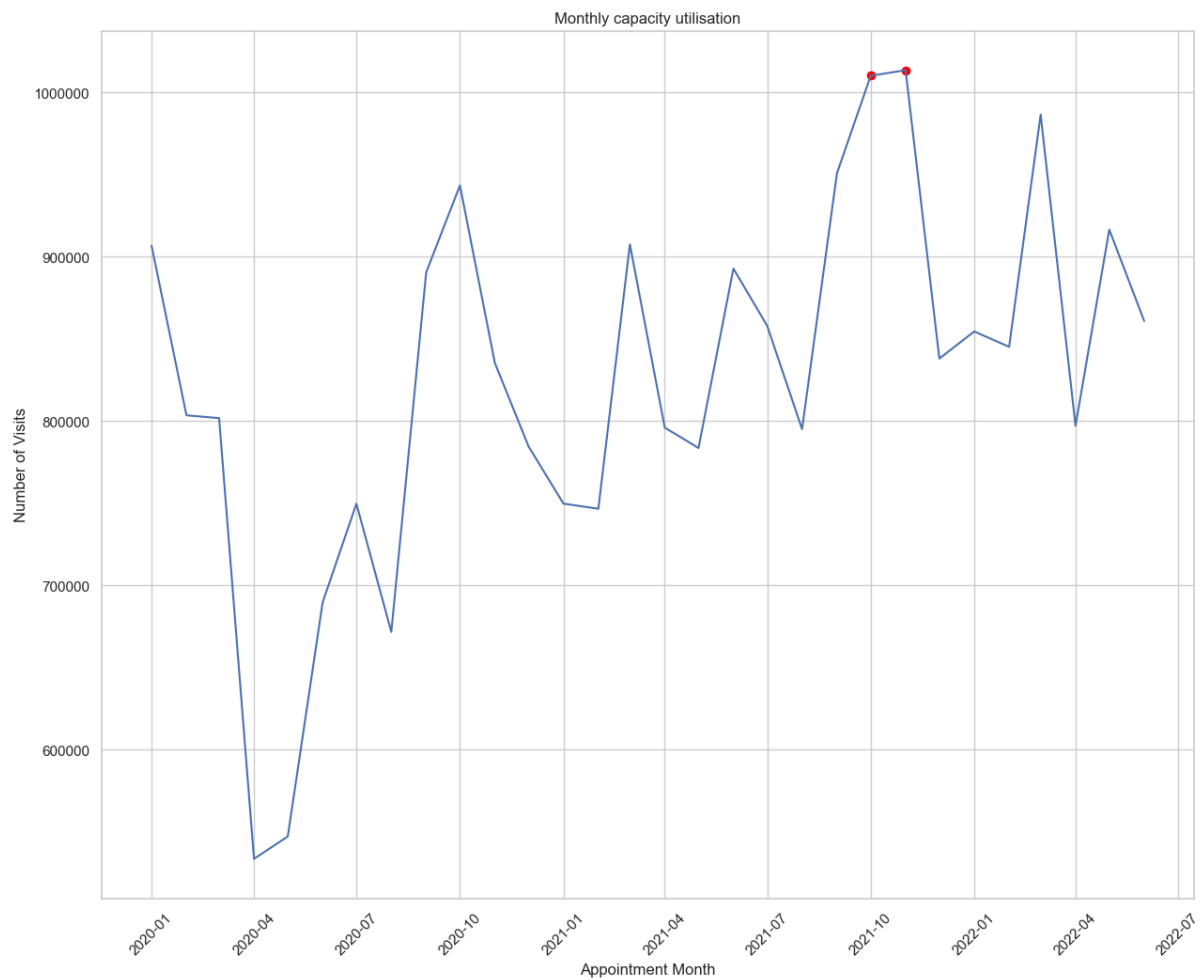


Fig. 1

Monthly capacity utilisation

Shows no months surpassed the daily appointment threshold of 1,200,000, but came close to between October 2021 to January 2022.

4.

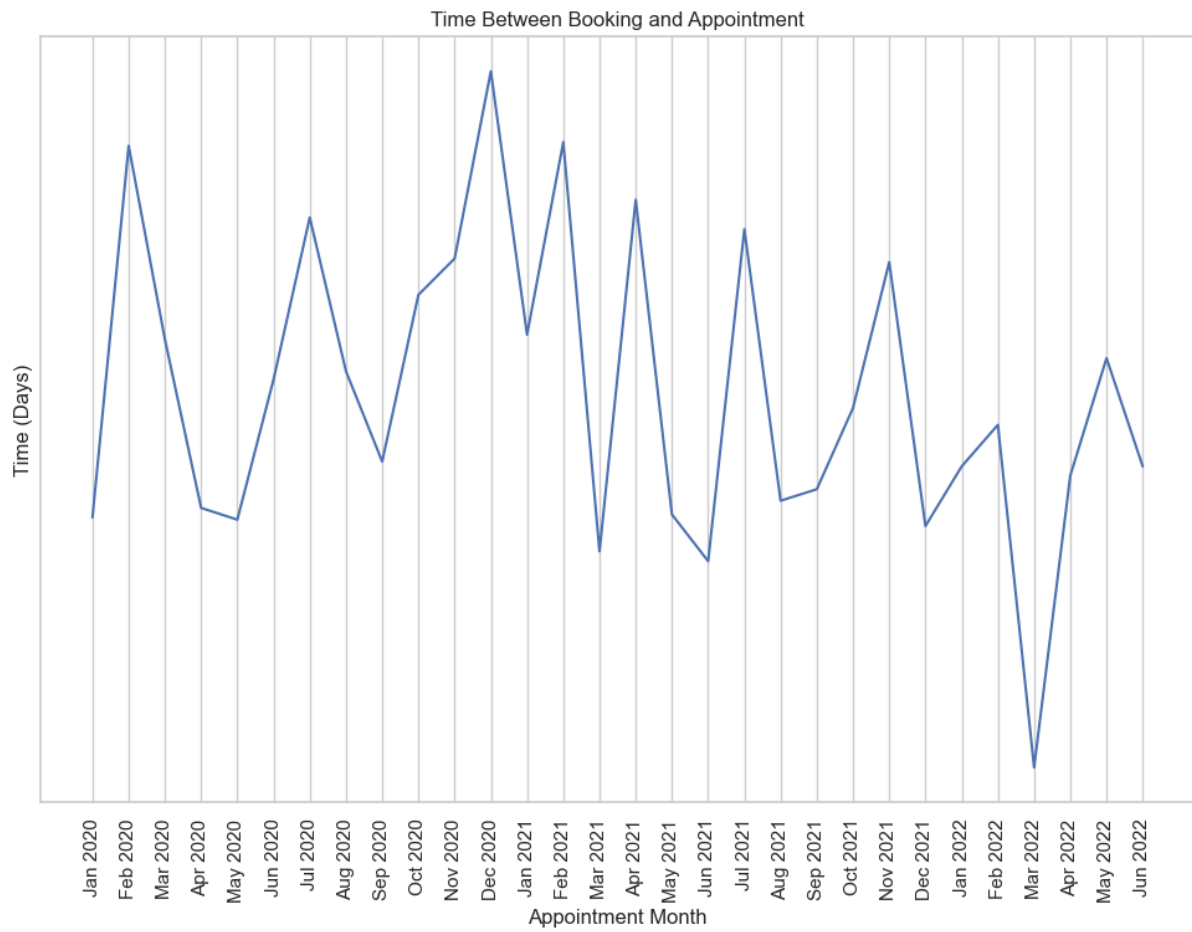


Fig 1

Time between booking and appointments

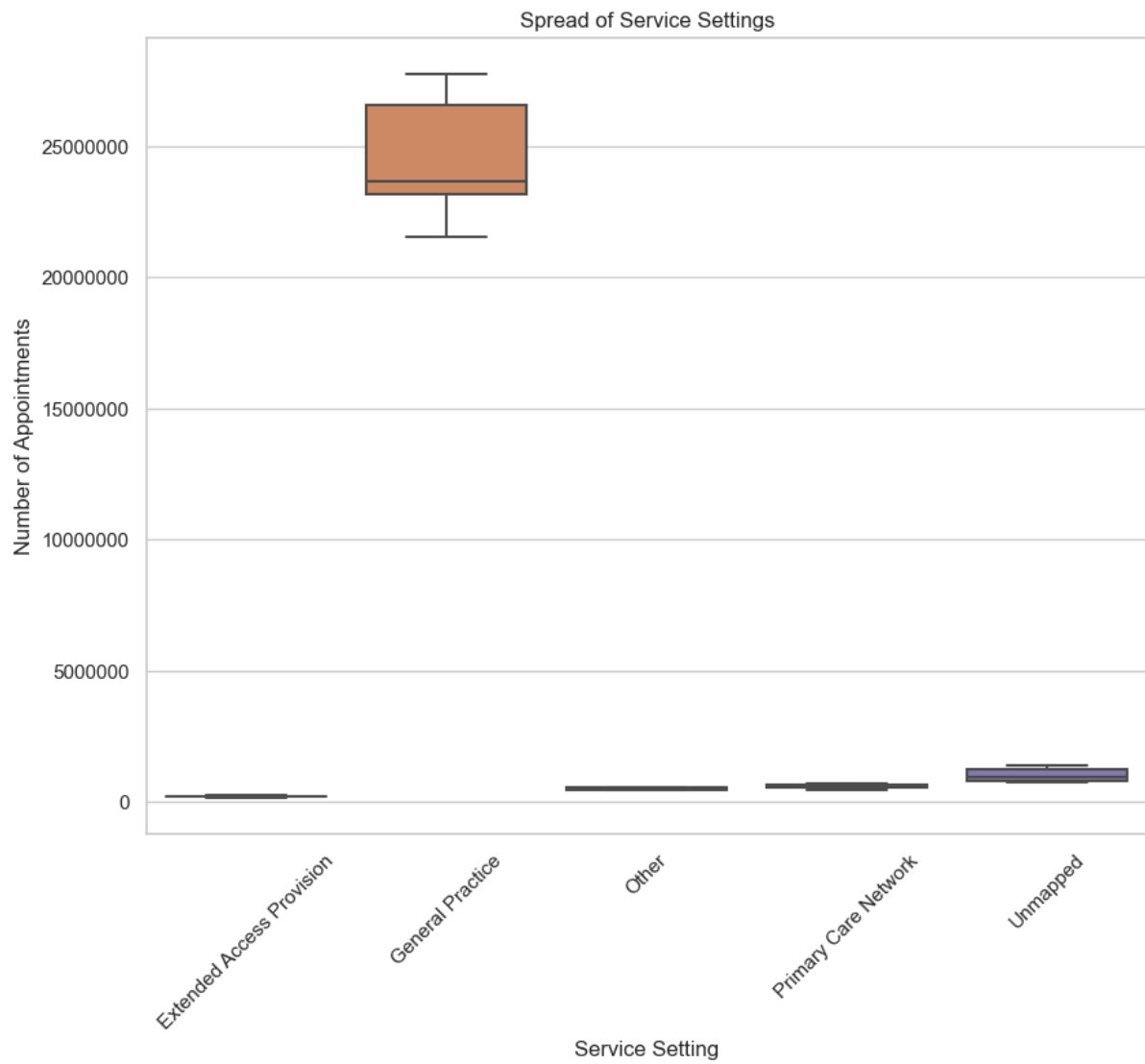


Fig. 1

Spread of Service Settings

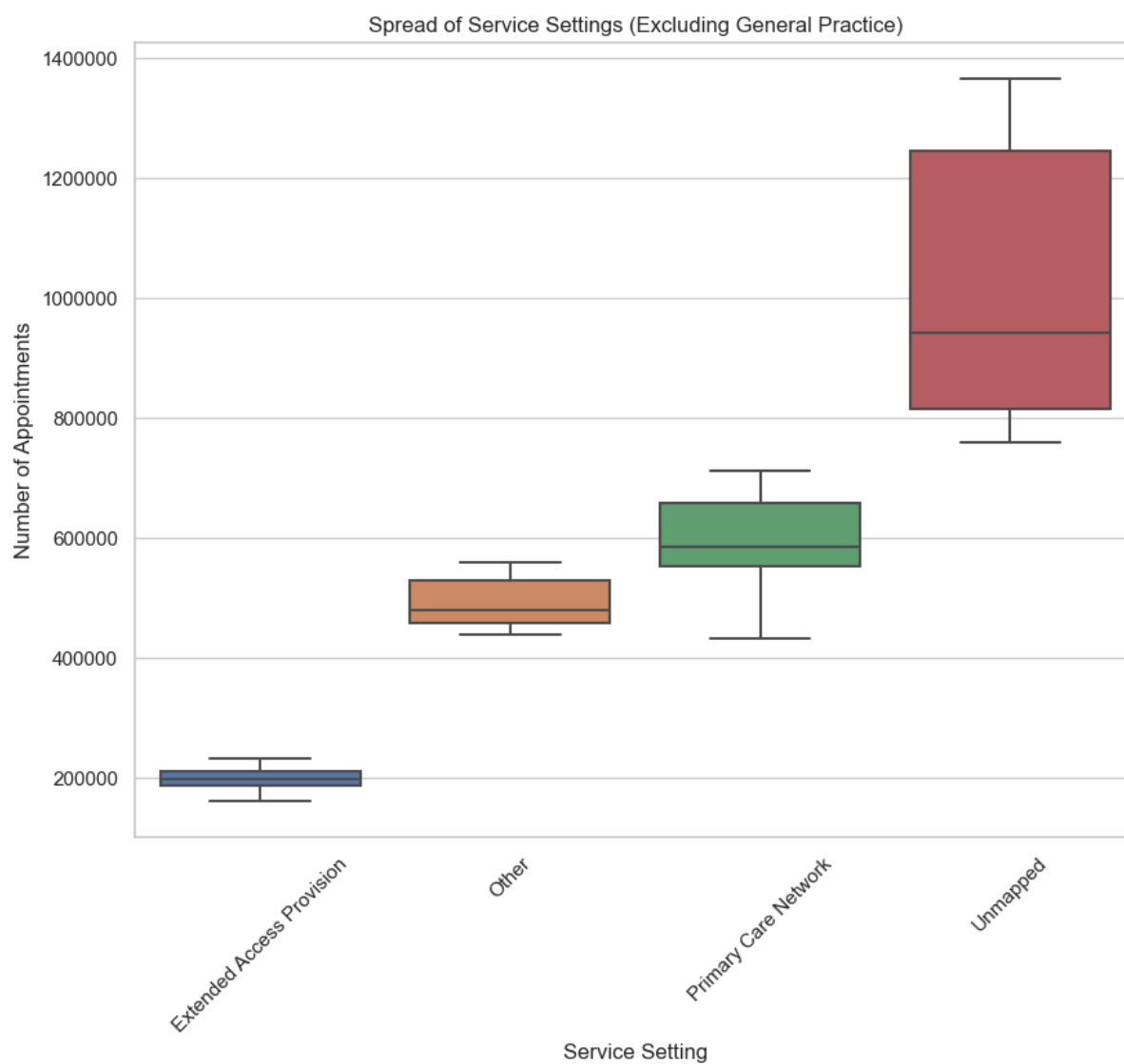


Fig 2.

Spread of Service Settings (Excluding General Practice)