

# NHS Key Insights

TECHNICAL REPORT DOCUMENTING AND EXPLAINING  
APPROACH AND INSIGHTS

DAVID GANDARY

## Background/context of the business scenario

- The government which controls the NHS a publicly funded healthcare system in England, needs a data driven approach to come to a decision on how best to tackle the issue of budget allotment.
- The problem faced is how best to allocate budgets with an overall question of whether the NHS should expand its capacity or make better utilisation of available resources.
- The two initial questions that have been asked by the NHS are:
  - Has there been adequate staff and capacity in the networks?
  - What was the actual utilisation of resources?
- Additional context has raised an argument that reducing or eliminating missed appointments would be beneficial financially as well as socially. (This is a key point that can be explored further.)

## Analytical Approach

The data analysis begins with the importation of relevant data files into DataFrames, followed by a thorough examination of data accuracy and cleanliness using methods such as `shape()` and `isnull().sum()`. The absence of missing values in the datasets is confirmed, ensuring their suitability for analysis.

- To gain an overview of the datasets, tables are generated to answer questions about the number of service settings, context types, national categories, and appointment statuses (**Appendices: A, B, C, D**). Key statistics, such as:
  - Total number of locations (817,394)
  - unique locations (106),
- These values are obtained through methods like `count()` and `nunique()`.

Further analysis focuses on identifying the most popular location, with **NHS North West London** standing out.

- The service setting that reported the most appointments for this location is determined to be General Practice.
  - Specification was required to ensure that data was being taken exclusively from the North West London location, which needed the unique location code and the required dates were also specified and these were combined to create a new variable.
  - The specified variable was then used to create a new subset by grouping the service setting column with the appointment date column.

- The analysis delves into specific months, revealing that November had the highest number of appointments. **(Appendix E)**
  - Important to note that the number of appointments per month ==sum of [count\_of\_appointments]. Therefore, this column was grouped by the [appointment\_date] columns which was split into its month and year using a datetime function.
- An evaluation of staffing adequacy involves assessing monthly staff utilization, displayed through a line graph. **(Appendix F)**
  - This had been approached by working out whether the NHS should start looking at increasing staff levels.
  - the average utilisation per month which was calculated by dividing the sum of the monthly appointments by 30 to get an estimate for the daily number of appointments, which was then divided by 1,200,000. The NHS had provided a figure of an average of 1,200,000 appointments per day being used for planning purposes, which was used as a guideline for maximum capacity.
- A boxplot visualization, excluding General Practice, investigates the spread of data for various service settings. An outlier in the Extended Access Provision boxplot prompts comments on the accuracy and reliability of the data. **(Appendix G)**
- The analysis extends to a Twitter dataset, aiming to identify top trending hashtags. Insights are derived from retweet and favourite counts, providing a gauge of consensus. **(Appendix H)**
  - Stakeholders can leverage this data to understand the perception of the NHS and act accordingly, especially with regards to critical tweets that gain a lot of engagement.
  - The top 30 value counts for hashtags reveal that #healthcare is significantly more prevalent than others (785), raising considerations about its broad nature.
  - Appendix containing relevant visualizations support the findings and conclusions throughout the analysis.

## Visualisation and insights

The analysis incorporates various visualizations, primarily utilizing line plots and bar plots to showcase trends in the summation of records per month concerning service settings, context types, and national categories from August 2021 to June 2022. Legends are employed to distinguish categories, with distinct colours aiding clarity. General Practice emerges as the dominant service setting throughout, evidenced by a sharp decrease in appointments from November to December and March to April. Unmapped services exhibit a steady decline from October to January. **(Appendix I)**

Bar plots, delineating seasonal trends into summer, autumn, winter, and spring, provide additional insights. While the bar plot method may pose challenges in distinguishing closely valued bins, it is preferred over line plots to prevent overlapping issues. The overarching analysis reveals General Practice as consistently prevalent, implying its prioritization for detailed performance and well-being assessments. **(Appendices: J , K, L)**

Notably, seasonal data hints at an increased number of Primary Care Networks, promoting proactive, personalized, and coordinated health and social care for enhanced accessibility. This strategic utilization of resources aligns with the objective of optimizing patient care closer to their homes.

Refinements to visualizations include plot labels and titles for enhanced comprehension. On the x-axis, tick labels transition from datetime to string data types, ensuring clear categorical month distinctions. These adjustments aim to provide readers with a comprehensive and accurate understanding of the presented data. The emphasis on General Practice's sustained prevalence suggests its ongoing significance, while the consideration of Primary Care Networks underscores a strategic approach to resource utilization and patient care improvement.

## Patterns and predictions

- The analysis reveals valuable insights into healthcare resource utilization and appointment trends. During the 'winter,' there is a notable decrease in utilization capacity (0.7), signalling to stakeholders that there is room for additional staff in busier winter periods. **(Appendix F)**
- The highest summation of appointments in October corresponds with the highest count of missed appointments, potentially indicating a capacity issue (0.8). Patients may have been deterred by the overwhelming number of appointments, contributing to the increase in missed appointments. Reminders provide updates to patients that have booked appointments. This could be in the form of SMS and phone calls. **(Appendix M)**
- An intriguing anomaly occurs in October, where 'Other Practice Staff' records the highest sum of appointments, diverging from the usual dominance of General Practitioners. This anomaly is attributed to data entry errors grouping various occupations under 'Other Practice Staff,' which could provide contextual understanding for the deviation. **(Appendix I)**
- Examining resource allocation, the line plot **(Appendix N)** illustrates changes in appointment types, with face-to-face appointments peaking in October 2021, coinciding with the month's highest utilization and appointment sum. This potentially signifies a misallocation of resources, as a month with such high appointment numbers might have benefitted from an increased focus on non-face-to-face appointments for more efficient resource utilization. Importantly, the context of the COVID Omicron variant's emergence during this period emphasises the need for optimized resource allocation during heightened healthcare demands.

## References

### Bibliography

- Gov.uk. (2021). COVID-19 confirmed deaths in England to 31 October 2021 report. Available at: <https://www.gov.uk/government/publications/covid-19-reported-sars-cov-2-deaths-in-england/covid-19-confirmed-deaths-in-england-to-31-october-2021-report> (Accessed: February 2024).
- W3Schools. (n.d.). pandas DataFrame nunique() Method. Available at: [https://www.w3schools.com/python/pandas/ref\\_df\\_nunique.asp#:~:text=Definition%20and%20Usage,unique%20values%20for%20each%20column](https://www.w3schools.com/python/pandas/ref_df_nunique.asp#:~:text=Definition%20and%20Usage,unique%20values%20for%20each%20column). (Accessed: February 2024).
- pandas.pydata.org. (n.d.). pandas.DataFrame.nunique. Available at: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.nunique.html> (Accessed: February 2024).
- British Foreign Policy Group. (2020). COVID-19 Timeline. Available at: <https://bfpg.co.uk/2020/04/covid-19-timeline/> (Accessed: February 2024).
- British Medical Association. (n.d.). Primary Care Networks (PCNs). Available at: <https://www.bma.org.uk/advice-and-support/gp-practices/primary-care-networks/primary-care-networks-pcns#:~:text=Read%20our%20guidance%20for%20clinical,BMA%20services%20to%20guide%20you.&text=PCNs%20are%20groups%20of%20GP,services%20to%20the%20local%20population>. (Accessed: February 2024).

## Appendices

### Appendix A

Table demonstrating the total number of settings separated by categories detailing the different provisions of care in which the appointment was delivered, 'using national categories' data.

Service Setting	Count of appointments
General Practice	359274
Primary Care Network	183790
Other	138789
Extended Access Provision	108122
Unmapped	27419

### Appendix B

Table demonstrating the National Categories that are subdivided into three 'Context Types', using 'national categories' data.

Context Type	Count of appointments
Care Related Encounter	700481
Inconsistent Mapping	89494
Unmapped	27419

### Appendix C

Table which refers to each context type where each appointment type is contextualised, using 'national categories' data.

National Category	Count of appointments
Inconsistent Mapping	89494
General Consultation Routine	89329
General Consultation Acute	84874
Planned Clinics	76429
Clinical Triage	74539
Planned Clinical Procedure	59631
Structured Medication Review	44467
Service provided by organisation external to the practice	43095
Home Visit	41850
Unplanned Clinical Activity	40415

Patient contact during Care Home Round	28795
Unmapped	27419
Care Home Visit	26644
Social Prescribing Service	26492
Care Home Needs Assessment & Personalised Care and Support Planning	23505
Non-contractual chargeable work	20896
Walk-in	14179
Group Consultation and Group Education	5341

## Appendix D

Table which describes the nature of the appointment in terms of being available, booked, been attended or had been cancelled. DNA refers to 'Did Not Attend', using 'national categories' data.

Appointment Status	Count of Appointments
Attended	232137
Unknown	201324
DNA	163360

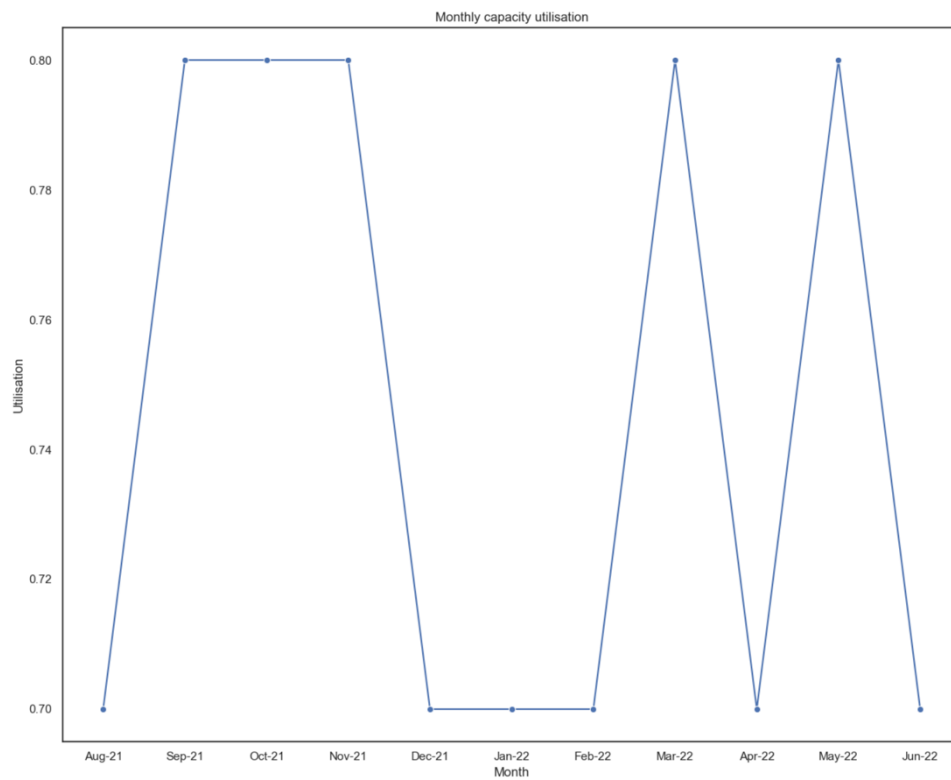
## Appendix E

Table which depicts the months and years and their corresponding sum of counts of appointments, using national categories data

Appointment Year	Appointment Month	Sum of Count of Appointments
2021	11	30405070
2021	10	30303834
2022	3	29595038
2021	9	28522501
2022	5	27495508
2022	6	25828078
2022	1	25635474
2022	2	25355260
2021	12	25140776
2022	4	23913060
2021	8	23852171

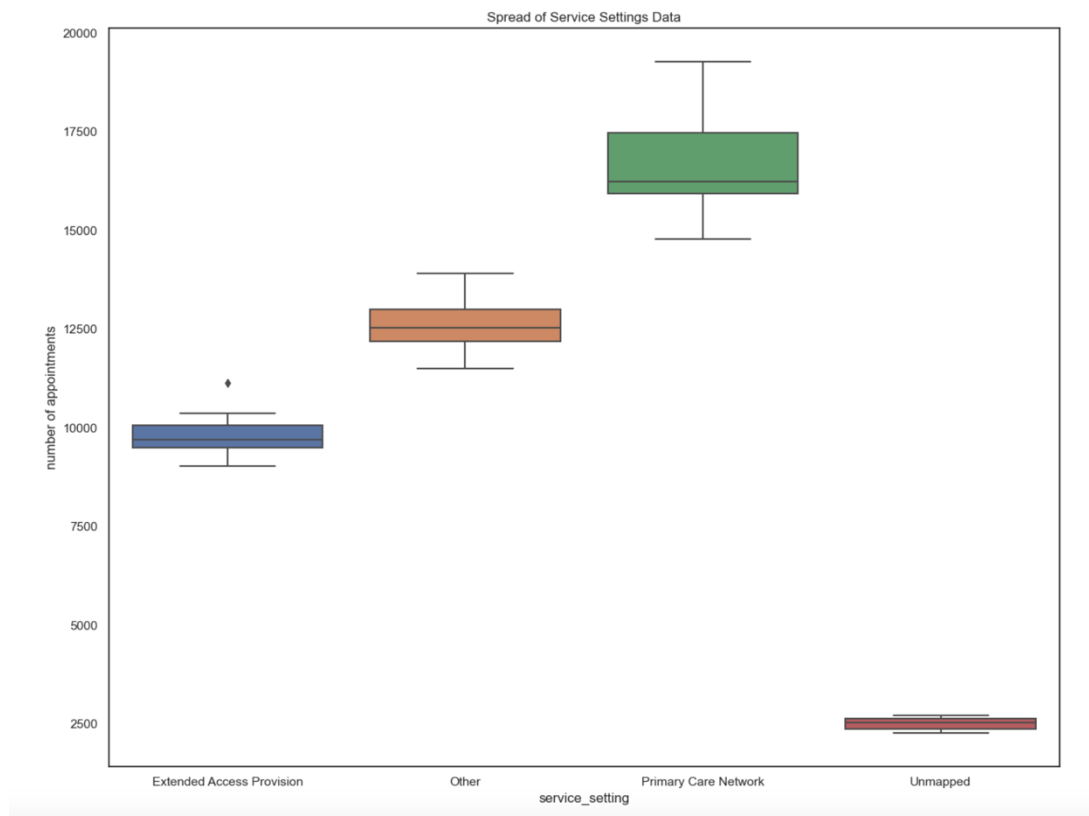
## Appendix F

Line graph which outlines the average utilisation per month.



## Appendix G

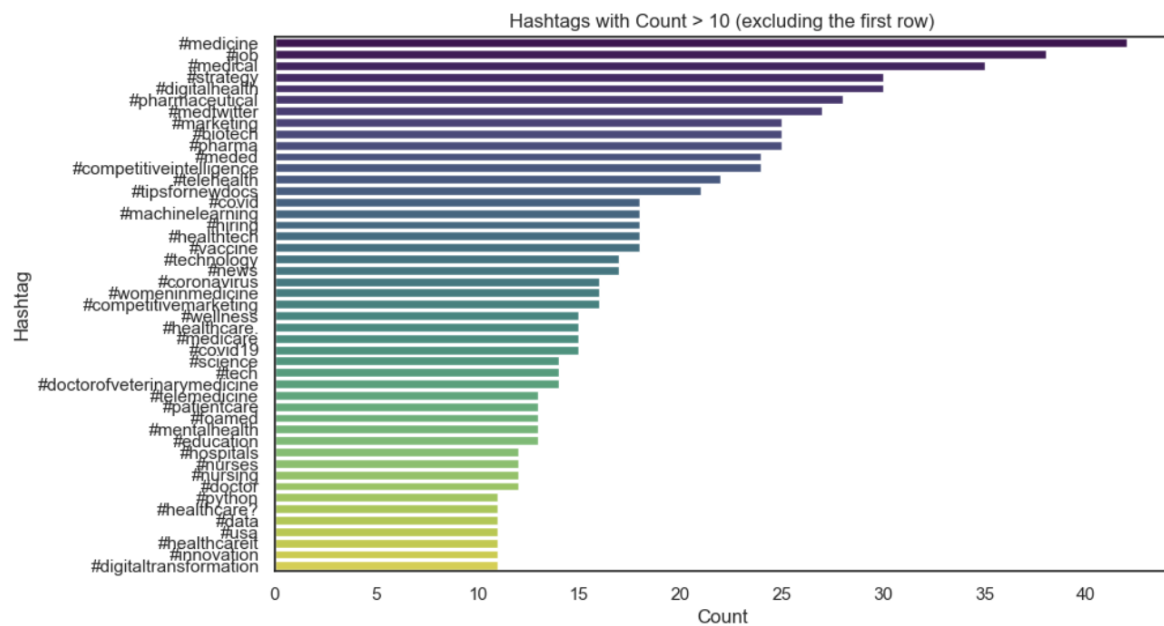
A boxplot visualisation which has the role of investigating the spread of service settings.





## Appendix H

Bar plot visualisation shows a similar output where the value counts for the top 28 hashtags.



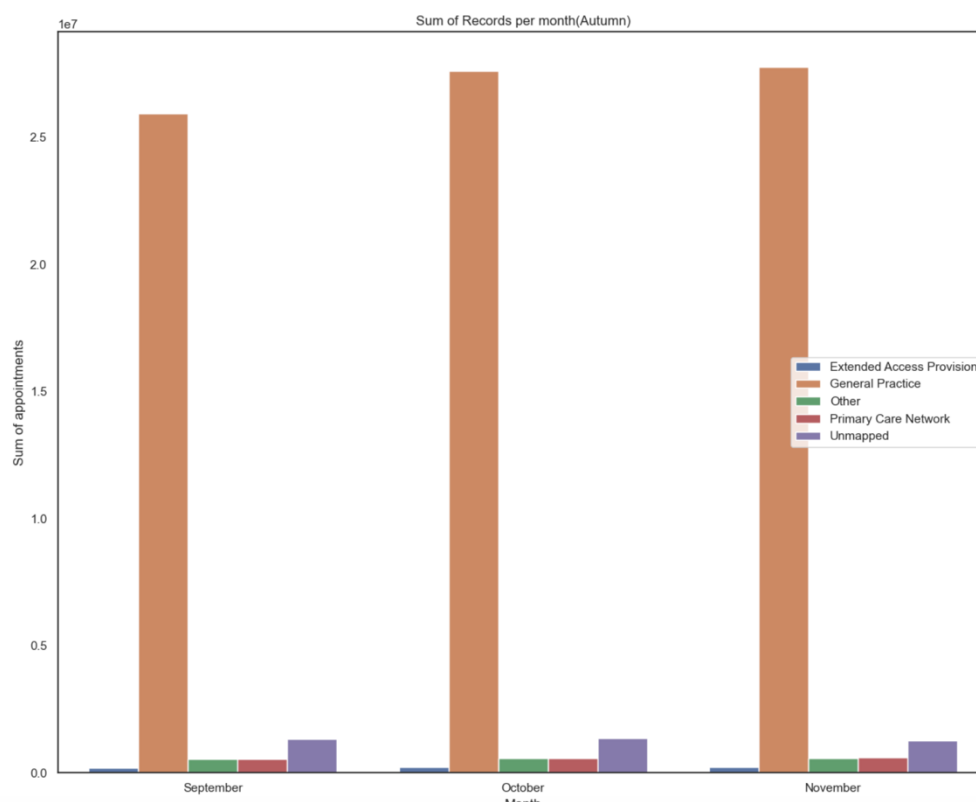
## Appendix I

Line plot outlines how the healthcare professional types differ over time.



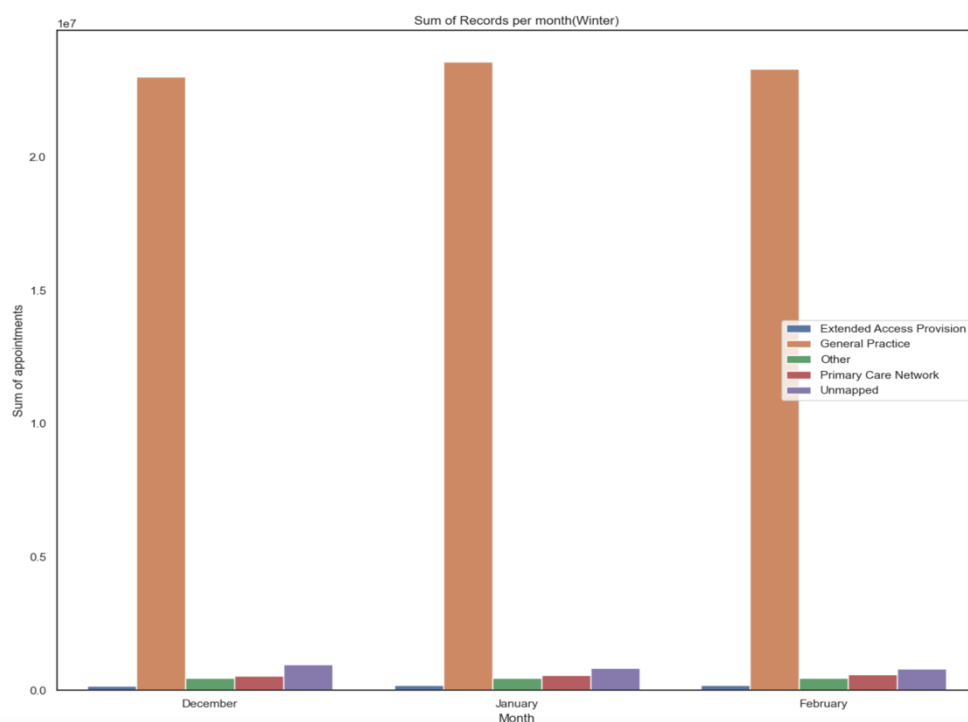
## Appendix J

Bar plot depicts the sum of records per month for the autumn season.



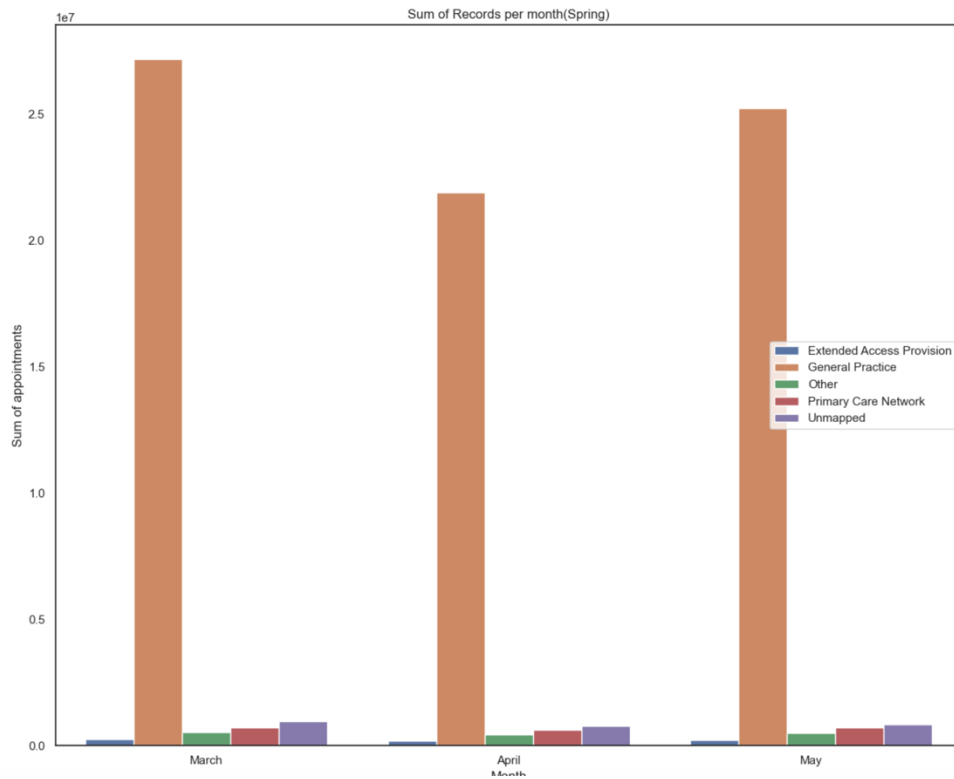
## Appendix K

Bar plot depicts the sum of records per month for the winter season.



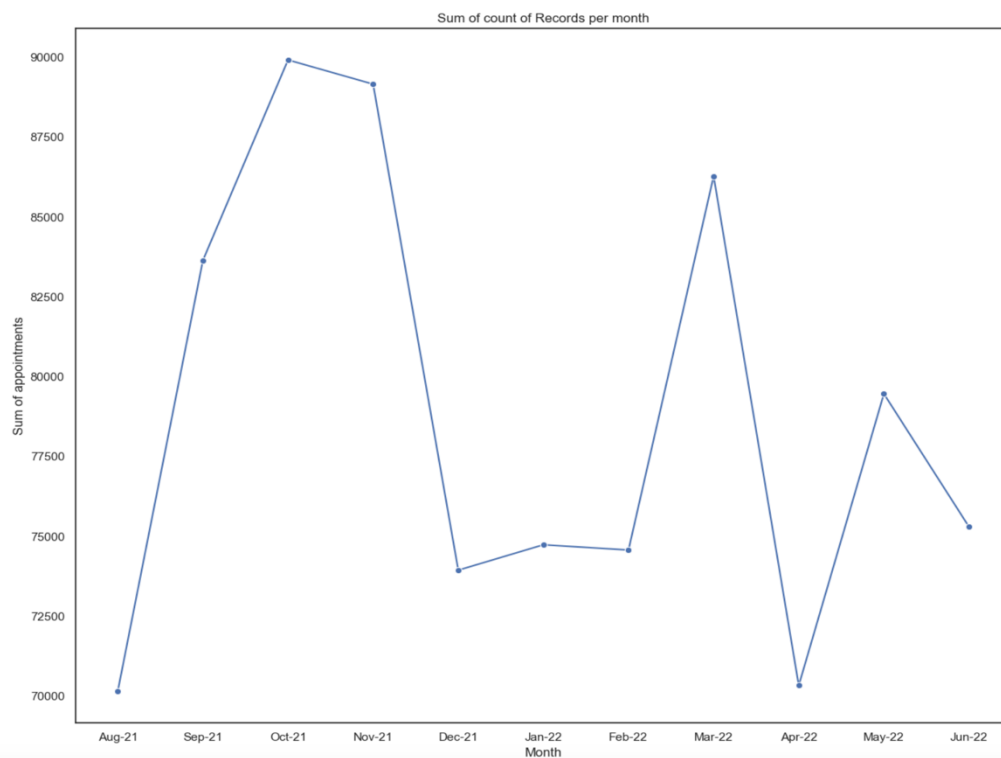
## Appendix L

Bar plot depicts the sum of records per month for the spring season.



## Appendix M

A line graph which outlines the sum of the count of records per month.



## Appendix N

Line plot which highlights changes in terms of appointment type and the busiest months.

