



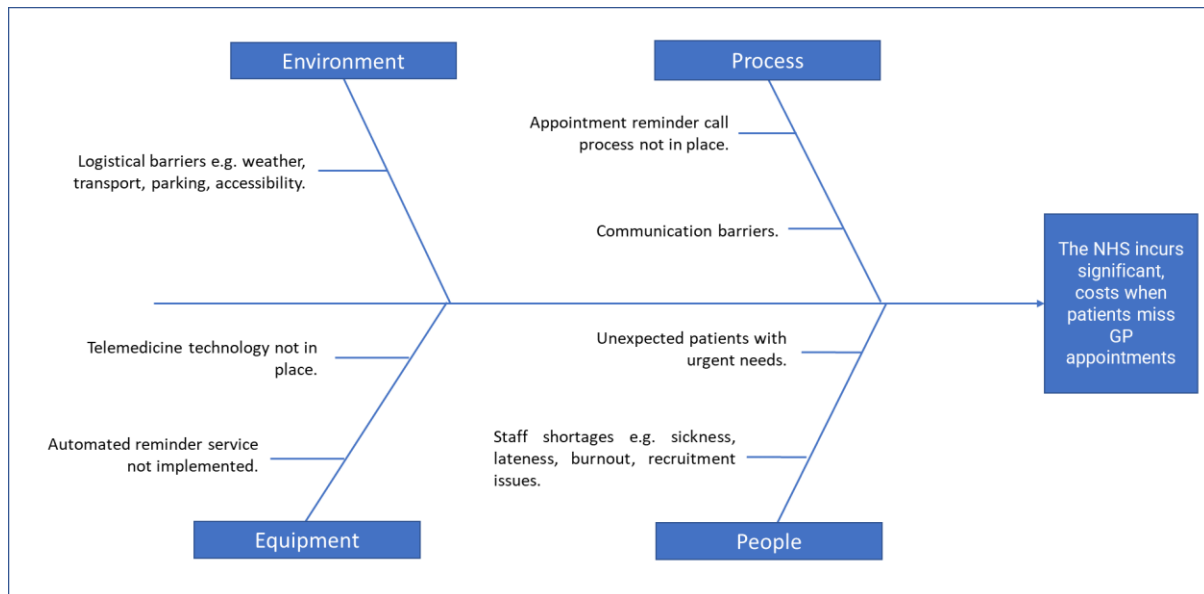
NHS Data Analysis

NHS Data Analytics Project – February 2022

1. Background

The Primary Care team for the Northwest London ICB, NHS NW London, commissioned this project to better understand potential reasons for missed GP appointments, which lead to significant, potentially avoidable, costs for the NHS.

Figure 1. Fishbone diagram showing potential causes of missed GP appointments.



This report includes seasonal analysis of appointments in NHS NW London, and a comparison to the whole NHS for factors impacting appointments, including HCP type, appointment mode, and time between bookings. It will also provide an estimation of appointment utilisation rates compared to the whole NHS, and an analysis of Twitter data related to healthcare.

2. Analytical Approach

Metadata for NHS files indicated data was clean. Checks for unusual data, spelling errors, duplicates, and data type found no issues. No metadata provided for tweets.csv, but checks detected no issues.

Data import into Jupyter Notebook:

- Code written in line with PEP8 - short, descriptive function/variable names, concise syntax, and insightful commentary.
- Raw files added into Jupyter directory and Python libraries/packages imported.
- DataFrames created using read() for actual_duration (ad), appointments_regional (ar), national_categories (nc), and tweets. For ad, certain columns imported with usecols.
- DataFrames checked for correct import; other methods used to explore data, e.g., shape(), info(), and isna().

Initial Data Exploration

- Date columns transformed using datetime function and strftime() as required. Date ranges determined with min() and max().
- Appointments by month, ICB, service setting, context type, national category, appointment status, HCP type, appointment mode, time from booking, and duration assessed.
- Methods/functions include value_counts(), groupby(), reset_index(), and sort_values().
- User defined function (calc_percent) created to calculate appointment percentage.
- Exploratory visualisations created using visual design principles.

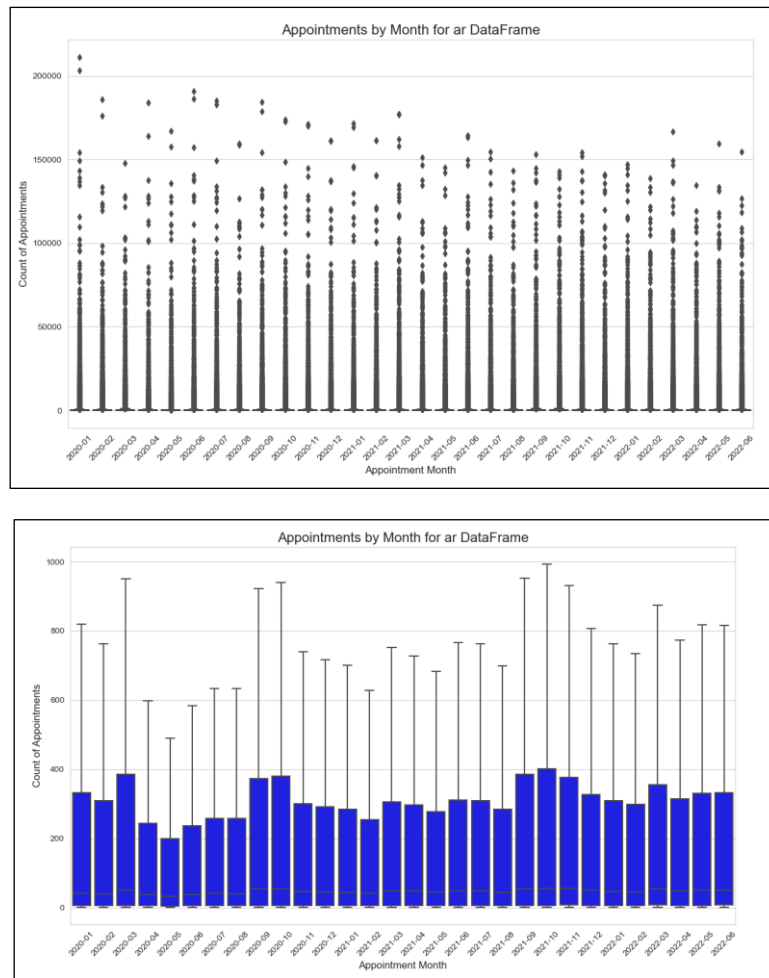
Outliers and Data Anomalies

- Outliers identified for ar, ad, and nc DataFrames using boxplots. Data without outliers visualised using showfliers argument.
- Outliers not removed as datasets would have decreased significantly (Figure 2).
- DataFrames also contained Unknown, Unknown / Data Quality, Unmapped, or Inconsistent Mapping data. Cleaned DataFrames created, with records removed using drop().

Social Media Data.

- A for loop used to cycle through tweet text looking for hashtags. Results appended into a list and a bar plot created of top hashtags.
- Tweets searched for terms e.g., 'job', 'career', 'nhs', and 'appointment'.

Figure 2: Boxplot showing outliers for the ar DataFrame (top). Boxplot following outlier removal (bottom).



Main Analysis

- Subsets of cleaned DataFrames created for NHS NWL between December 2021 and June 2022 (remove COVID-19 impact).
- **nc_nwl DataFrame:** seasonal appointment trends visualised for service settings using line charts.
- **All DataFrames:** bar charts created comparing appointment percentages in NHS NWL to whole NHS for status, service setting, HCP type, mode, time between booking, and duration. Used `groupby()`, the user defined function, and `merge()`.
- **ar_nwl DataFrame:** line charts created to see impact of HCP type, appointment mode, and time between booking on appointment status.
- **Main ar and ar_nwl DataFrames:** daily appointment utilisation rates compared for NHS NWL and whole NHS.
- **Main ad and ad_nwl DataFrames:** percentage of appointments in NHS NWL and whole NHS above and below target of 10 minutes compared.
- **Uncleaned ar and ad DataFrames:** merged to check for correlation between appointment duration and HCP type.

3. Visualisations and Insights

Visualisations created using basic visual design principles, concerning chart type, colour, size, and layout. Bar charts were mainly used for comparisons, whereas line charts were used for time series visualisations.

Insights from initial data exploration:

- Impact of COVID-19 lockdowns on monthly appointments.
- Over 90% of appointments attended, and most booked on the same day.
- Most appointments between 1-10 minutes, and face-to-face and telephone most common.
- Most appointments conducted in General Practice, and a similar proportion conducted by GPs and Other Practice staff.

Figure 3: Monthly appointment trends across all ICBs, January 2020 – June 2022.

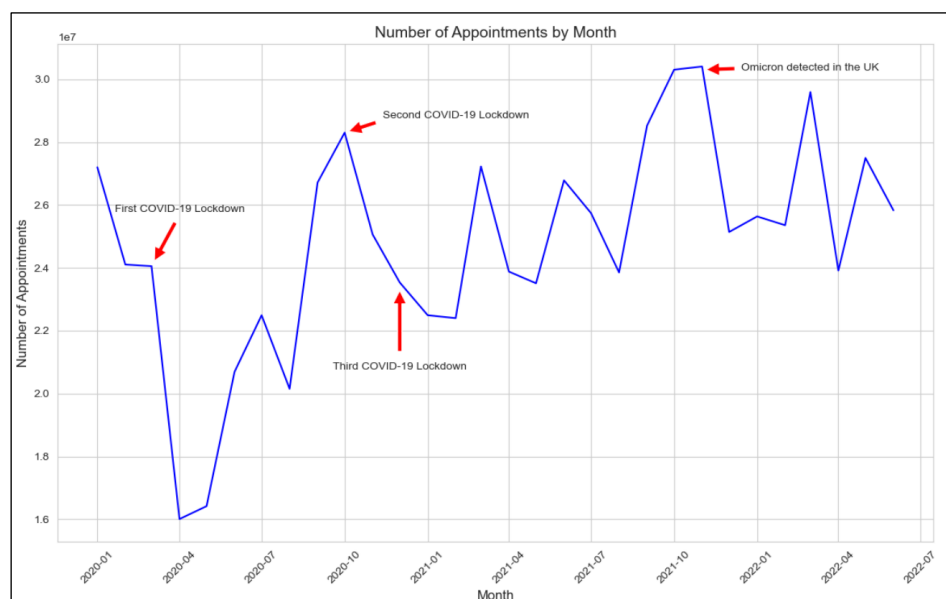


Figure 4: Appointments by status.

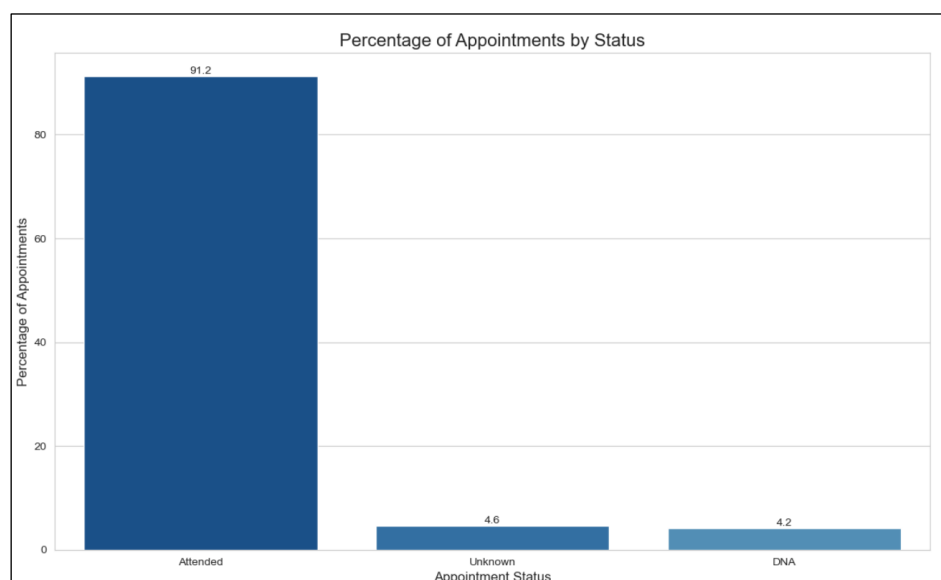


Figure 5: Appointments by time between booking.

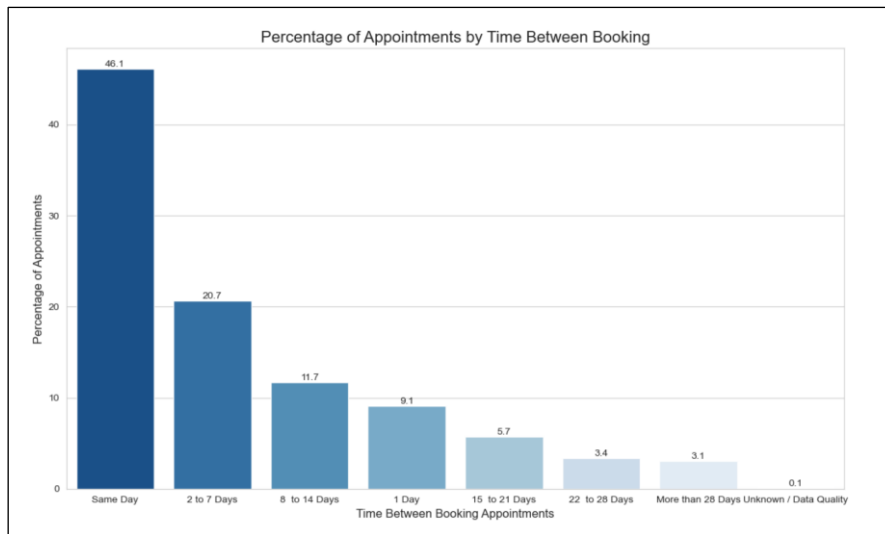


Figure 6: Appointments by duration.

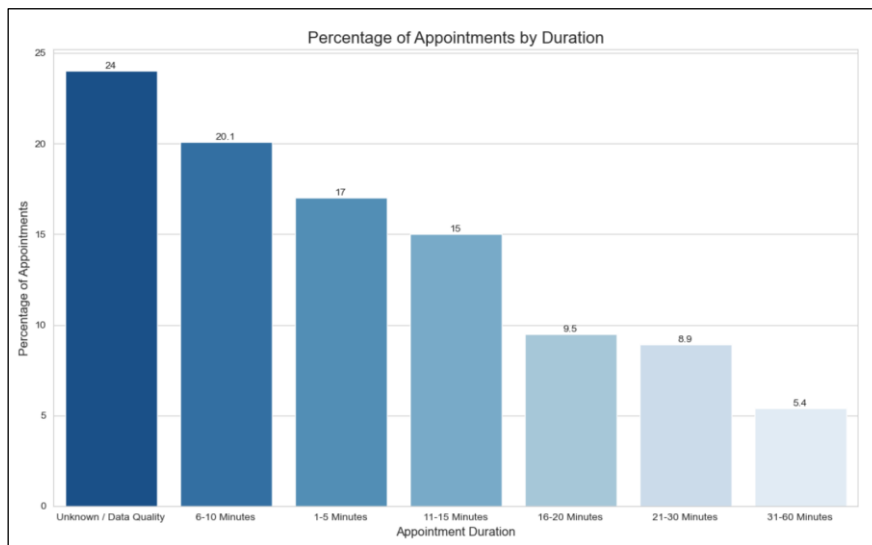


Figure 7: Appointments by mode.

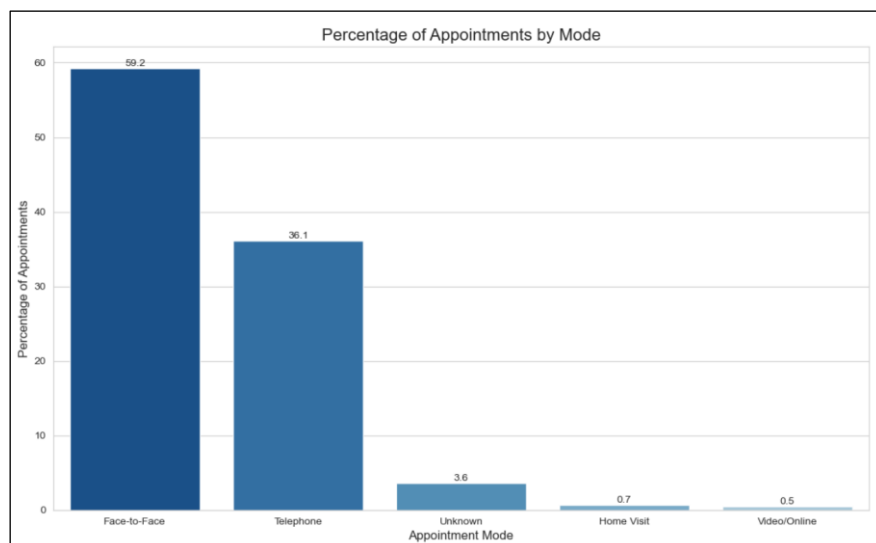


Figure 8: Appointments by service setting.

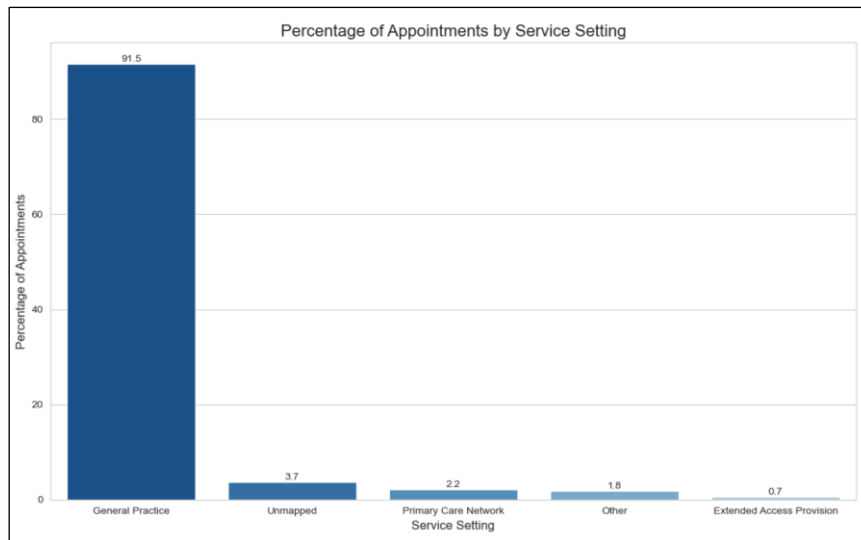
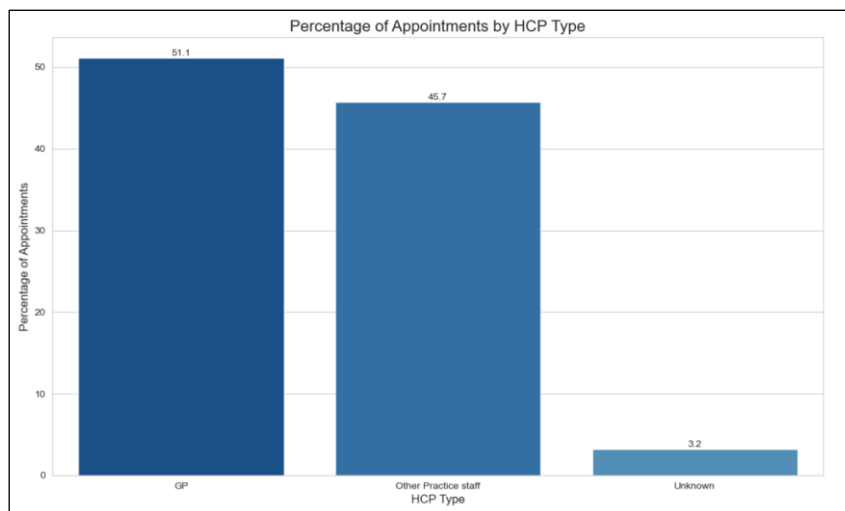


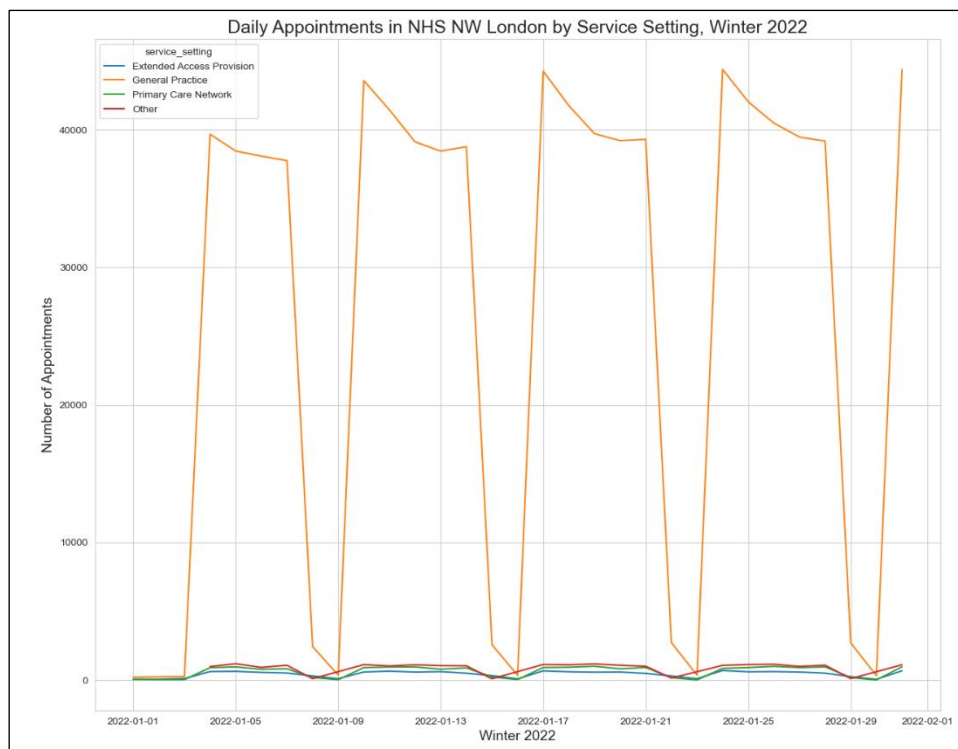
Figure 9: Appointments by HCP type.



Insights from main analysis:

- **Seasonal appointment trends in NHS NWL:** line charts for daily appointments for Winter, Spring, and Summer 2022 showed similar trends.
 - More appointments at the start of the week for each season - due to demand after the weekend.
 - For Winter 2022, weekly appointments increased throughout the month - due to demand due to COVID-19 and other winter illnesses.

Figure 10: Daily appointment trends in NHS NWL in Winter 2022 .



- NHS NWL vs NHS appointment trends:** NHS NWL trends generally comparable to whole NHS for status, service setting, and HCP type.
 - NHS NWL uses **more virtual/online appointments** (2.8% vs. 0.5%).
 - NHS NWL had fewer same day appointments - but **higher between 1-14 days and fewer over 15 days**.

Figure 11: Percentage of appointments by status, NWL vs. NHS.

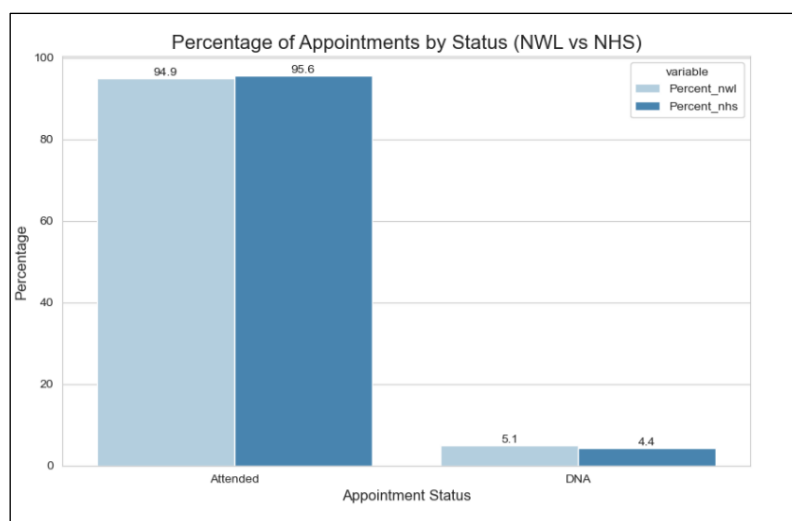


Figure 12: Percentage of appointments by HCP type, NWL vs. NHS

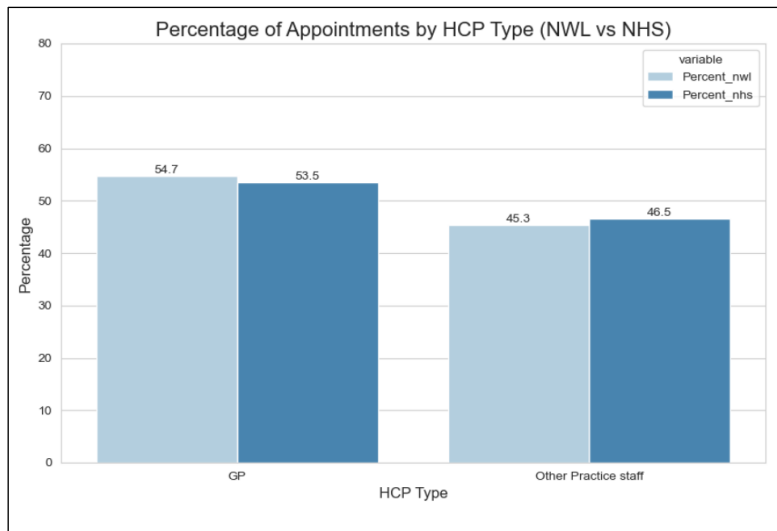


Figure 13: Percentage of appointments by mode, NWL vs. NHS.

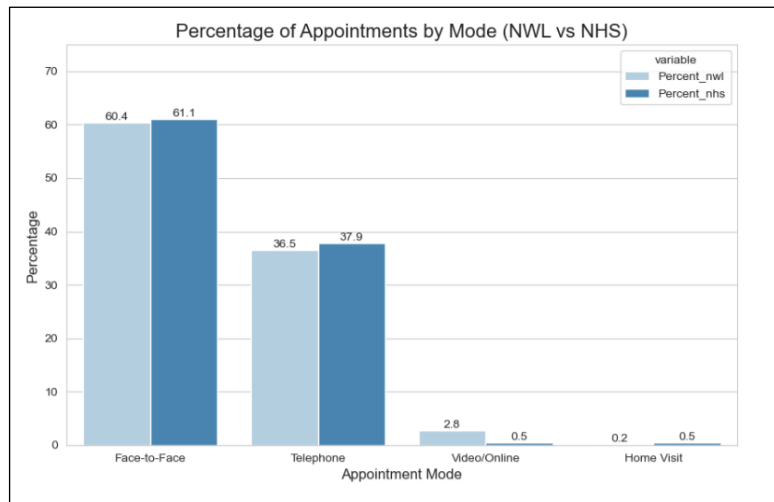
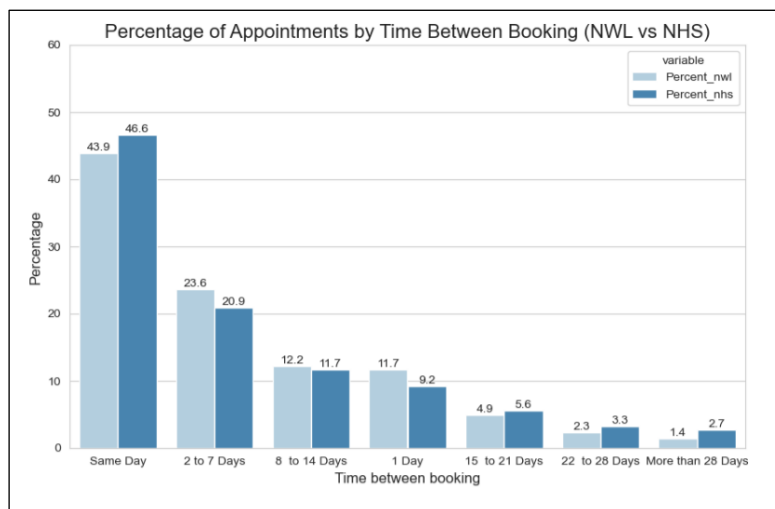


Figure 14: Percentage of appointments by time between booking, NWL vs. NHS.



- **Impact of HCP type, appointment mode, and time between booking on status:** Line charts showing the impact of appointment mode and time between booking on appointment status did not reveal any significant trends for NHS NWL.
 - However, charts for HCP type showed **that more unattended appointments were conducted by Other Practice staff.**

Figure 15: Impact of HCP type on attended (top) and DNA (bottom) appointments in NHS NWL.



- NHS NWL vs NHS daily appointment utilisation rates:** daily utilisation rates were calculated based on 1,200,000 daily appointments for the NHS and 48,000 for NHS NWL.
 - The resulting line chart shows that **NHS NWL had a consistently higher utilisation rate** over the 7-month period.
 - Further analysis of appointment duration shows that **NHS NWL has a higher percentage of appointments that fall within the 10-minute target.**

Figure 16: Daily appointment utilisation rates, NWL vs. NHS.

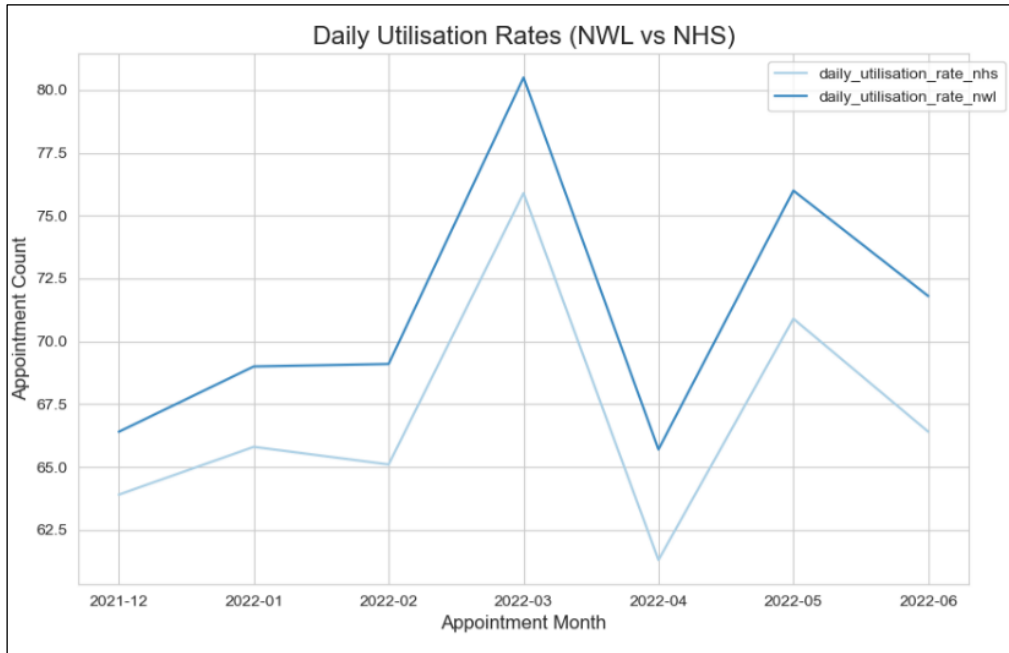
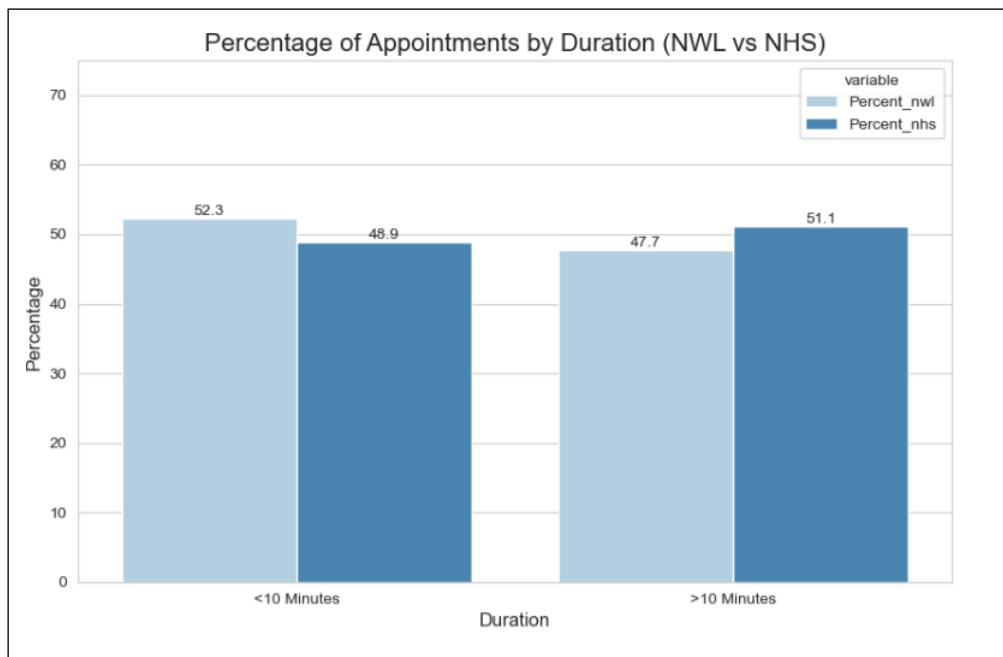
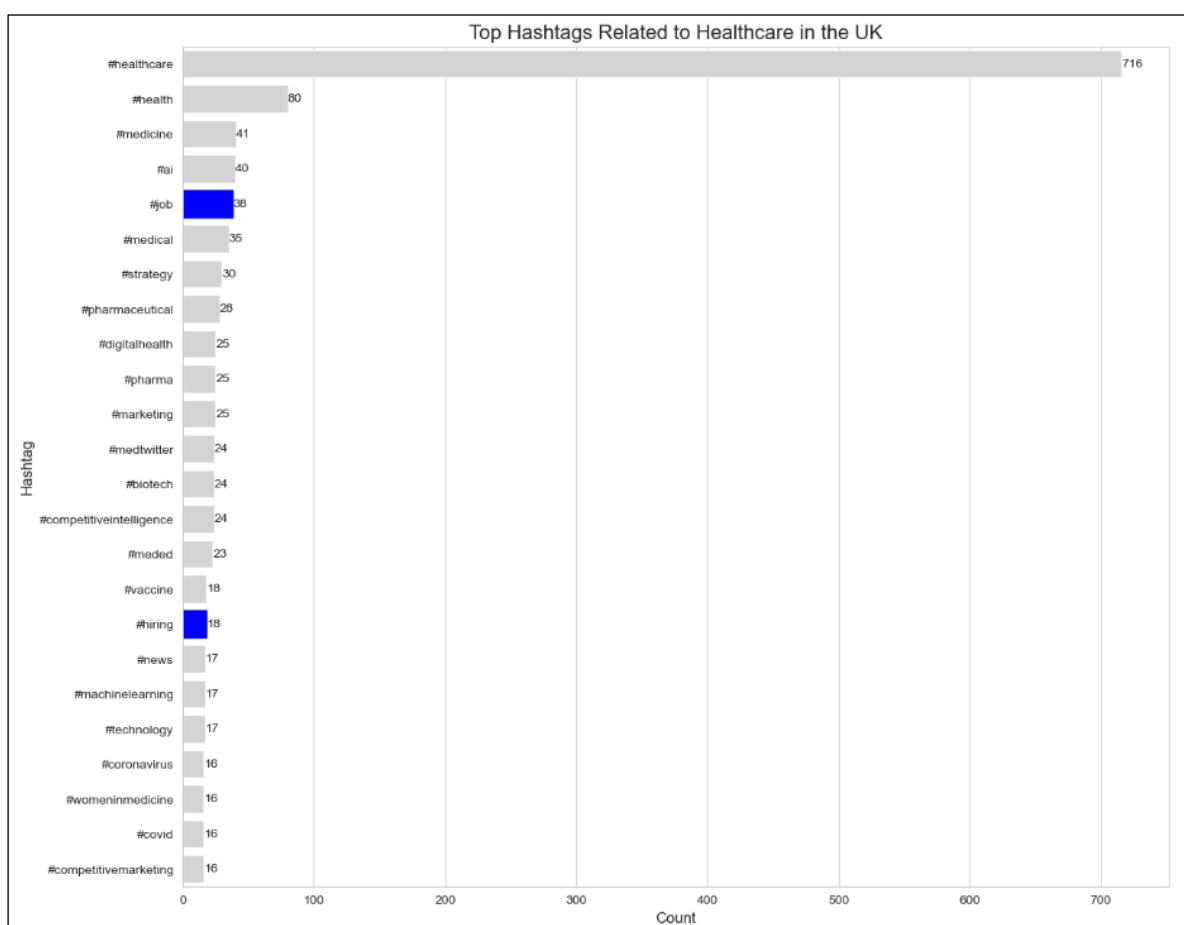


Figure 17: Appointment duration below and above 10 minutes, NWL vs. NHS.



- **Correlation between appointment duration and HCP type:** data limitations meant any relationship ar and ad DataFrames could not be explored. More information on appointment duration by HCP type required.
- **Insights from social media:**
 - Searching tweets containing 'nhs' or 'appointment' did not return insights to explain missed GP appointments.
 - There were posts related to jobs and hiring which could be indicative of staff shortages across healthcare.

Figure 18: Top hashtags related to healthcare in the UK, with those related to jobs highlighted.



4. Patterns and Predictions

- **Open weekend hubs or have longer weekday opening hours** to meet higher demand for appointments at the **start of the week**.
- Encourage **virtual appointments to remove barriers to care** and target groups with the required technology.
- NHS NWL performs well for booking times between appointments and should keep these below 15 days. Patients may forget or seek advice/treatment elsewhere. Implementation of an **automated reminder system** is recommended.
- More missed appointments involve Other Practice staff, likely due to less urgency. Implement an **automated service** to patients one week prior about planned attendance.
- NHS NWL has a **higher daily appointment utilisation rate** compared to the whole NHS. Reducing appointment duration and use of telemedicine could improve this further. Almost half of appointments are over the national 10-minute target.
- There is a shortage of healthcare workers due to Brexit and COVID-19. **Use social media to recruit**, as a wider reach could entice healthcare workers from abroad.
- **Areas for further exploration:** Additional data on attended appointments to analyse impact of HCP type and appointment mode on duration.
- **Data limitations:** There was no date for social media data, or what proportion of total tweets were related to healthcare to assess trend popularity.