- Report will focus on the main technical aspects to answer the key business questions and additional analysis.
- Approach was to first answer the set business questions and then explore any extra analysis.
- Additional analysis focused around identifying causes of unattended appointments:
    - The rational is unattended appointments incur a ‘significant’ cost to the NHS.

- First task was to better understand the data sets through the number of:
    - Locations
    - Service settings
    - Context types
    - National categories
    - Appointment status

Syntax used:
len(pd.unique(df[‘column_names’]))
df[‘column_names’].value_counts

- len function combined with pd.unique determines the number of unique values for each variable.
- value_counts provide the record count per each value in a column
- Output in appendix item: 1.a
- Additional part of understanding the data sets was to examine the date range.

Example syntax:
df['column_name’] = df['column_name'].astype('datetime64')

- Changed the data type of ar, nc & ar df to datetime64.
- Limitation of the data is the difference between when appointments were scheduled e.g. ad between 2021-12-01 and 2022-06-30 and nc between 2021-08-01 and 2022-06-30.

Additional Analysis

- Explored the nc database further to analyse the difference between the top location (NHS North West London ICB - W2U3Z) and bottom location (NHS Greater Manchester ICB - 00V) for count of records.
- Interesting comparison as both sub sections of major cities.

Example syntax:
nc[nc['sub_icb_location_name']=='NHS North West London ICB - W2U3Z']

- There is a disparity between the sum of appointments and the record count in relation to unmapped services for both locations.
- Suggests poor data collection as you can't determine what this service is. Limits the impact to make notable insights regarding the use of services in each location.
- Appendix 2.a.b.

Monthly and Seasonal trends for service settings, context types and national categories:

- Three new DataFrames created from the nc DataFrame to separate necessary information being: service setting, context type and national category.
- Rationale is to provide insight into monthly and season trends based on each of these variables via lineplots.

```
new_df = df.groupby(['appointment_month', 'column_variable])\
            .sum().reset_index()

sns.lineplot(x='appointment_month', y='count_of_appointments',
                hue='service_setting', data=nc_ss, ci=None)
```

- Limitation of the service setting lineplot is the significant difference between the General Practice appointments compared to other service settings, creates a large amount of white space on the graph.
- Trends identified regards General Practice:
    - Sharp rise in appointments during autumn months – expected due to the change in seasons coming out of summer.
    - Followed by a significant fall at the beginning of winter (Nov 21 27767889 – Dec 21: 23008818). Winter sees a relative flattening

of the line which can be put down to the holiday period of Christmas and New Year.
    - Spring 2022 starts with a sharp rise in March followed by a sharp drop in, which can be credited to the bank holidays causing four days of no appointments - looking at the isolated apr_22 graph there is significant drop around these dates.
    - Appendix 3.a.b.c.
- A new DataFrame was created to exclude (!=) General Practice. Enabled a better insight into the other service settings.

```
nc_ex_gp = nc_ss[nc_ss['service_setting'] != 'General Practice']
```

- Key Insights:
    - Primary Care Network sees a steady increase in appointments over the analysed period – probably a result of coming out of Covid-19 where visits by medical staff is easier.

- Also split the nc_nc dataframe, due to too many lines and the graph being clustered.
- nc_nc dataframe split by national categories with a total of 10 million or more appointments and then those below.
    - Groupby() and sum() functions used to determine number of appointments.
    - .isin function used to split each dataframe, identifying each variable.
    - Rise in 'Patient contact during Care Home Round' & 'Structured Medication Review'. Linked to increase in Primary Care Network service – confirms theory of coming out of Covid-19
    - Appendix 4.a.b.c.d.

Top trending hashtags:

- After importing the tweets file (tweets = pd.read_csv('tweets.csv')), created a empty list and utilised a for loop to separate the hashtags.
    - Appendix:
- Converted list to a series using pd.series and then converted this series to a dataframe using pd.Dataframe.
    - Appendix 5.a.b.
- A new df created to filter even further for hashtags with a count > 10

- These hashtags are plotted on a barplot, with bars running horizontally due to the amount of variables.
- Limitation of this graph is twofold:
    - The sheer number of variables.
    - The top values being 'parent' hashtags, cover the broad topic of healthcare, which offers little insight.

- Therefore, created a new df to filter out the outliers in the previous chart.
- Removal of outliers achieved through calculating the lower and upper quartile range for the df and running a Boolean array.
    - Appendix 5.c, d & e

Utilisation of resources

- Created a new df filtering ar from August 2021 onwards, to bring it in line with the dates of the nc dataframe.
    - Appendix 6.a.
- A new df (ar_agg) created to group key columns by count of appointments.
    - Appendix: 6.b.
- Further df created to group each month by appointment count, from this utilisation could be calculated.
- NHS can accommodate 1,200,000 appointments a day.
    - Divide each month's count of appointment but 30 = avg daily total.
    - Avg daily total / 1,200,000 = utilisation.
    - Appendix: 6.c.d.
- As expected during the months where there is a fall in appointments there is also a decrease in the utilisation of services.
    - Lower number of appointments the increase in spare appointments therefore the lower utilisation.
    - More analysis required to investigate the utilisation of each service. Need info on maximum appointments each service setting can accommodate.

Extra Analysis: Causes of missed appointments

- Unattended appointments are a significant cost to the NHS.
- Created a new df to separate the DNA appointment status.

    - Appendix 7.a.
- Lineplot produced to examine trend of total DNA appointments over the time period.
    - As expected, trend follows the same pattern of total count of appointments. More appointments the more unattendances.
    - Appendix 7.b.
- Referring back to the time between book and appointment lineplot:
    - 2 to 7 days and 8 to 14 both experience peaks at the same time there is an increase in unattended appointments.
- Therefore, filtered the data to focus on DNA appointments grouped by each time between book and appointment.
    - Appendix 7.c.
    - Insight here is that 2 to 7 and 8 to 14 responsible for the most missed appointments – particularly in October 2021. This month is notable as it is when 'Other Practice Staff' exceeds GP appointments
- New df created to isolate 2 to 7 and 8 to 14 days to see which hcp type they use.
    - As predicted the hcp type with the most appointments for both 2 to 7 and 8 to 14 days' time between book and appointment is Other Practice staff.
    - Starting to build an understanding of the causes for unattended appointments.
    - Appendix 7.d.e.f.g.
- Lastly, I wanted to look at hcp type per DNA appointments, with the prediction Other Practice staff would be the main contributor.
    - Prediction was correct – the amount of DNA for Other Practice staff is significantly higher than GP & Unknown appointments.
    - Appendix 7.h.
- Conducted further analysis to determine the percentage of Other Practice staff DNA appointments to total Other Practice staff appointments.
    - Appendix:
    - Every month at least 6% of Other Practice staff appointments are unattended.
- Overall insights:
    - 2 to 7 days and 8 to 14 days appointments contribute significantly more than any other time book between appointment for unattended appointments.

    - This is likely because Other Practice staff is the most common type of appointment for this time frame – this hcp type has considerably more DNA appointments than General Practice appointments.

Recommendations:

- Deeper research into Other Practice staff
    - Be more specific to determine which are causing unattendances
- Conduct research into why patients do not attend appointments with a time between book and appointment of 2 to 7 days and 8 to 14 days.
    - Is it because its Other Practice staff or another reason?
- Provide data on staff levels
    - Impossible to make an accurate insight on staffing levels without the relevant information.
- Improve data collection as there are multiple issues as per the metadata:
    - DNA appointments not captured correctly.
    - HCP type was incorrectly extracted for some practices from Oct 2017 onwards – only GP extracted correctly.
    - Widespread variation in approach to appointment management between practices. This means that there are variations in data quality between practices.
    - High count of unmapped and unknown appointments.
- Refer to bottom of Notebook for in-depth recommendations.