1) Background of Business Scenario

The NHS is finding it challenging to efficiently manage appointment bookings and reducing missed appointments across various service settings. A major issue is the high rate of unattended appointments, which in turn leads to recourses being used inefficiently which in turn leads to extended waiting times for other patients. This report will use the (EDA) data analysis framework to interpret trends in appointment types, the time between booking appointments and illustrating the busiest NHS periods (Congress, 1965). By comprehending all these factors, the NHS can become more efficient in tackling these business problems.

2) Analytical Approach

The analytical approach began with uploading the CSV data sets into python using Panda's library. The CSV files were turned into Data Frames to load the data so that the data could be manipulated and analysed. Additionally meta data was also loaded to help interpret the other data files.

2.1) Data Cleaning:

The datetime was converted so that the data could be analysed in order of dates. The 'appointment_date' and the 'booking_date' columns were converted to help with calculating instances when looking for time between bookings and appointments. Figure 1 below shows the code used to chronologically change the datetime.

(Figure 1)

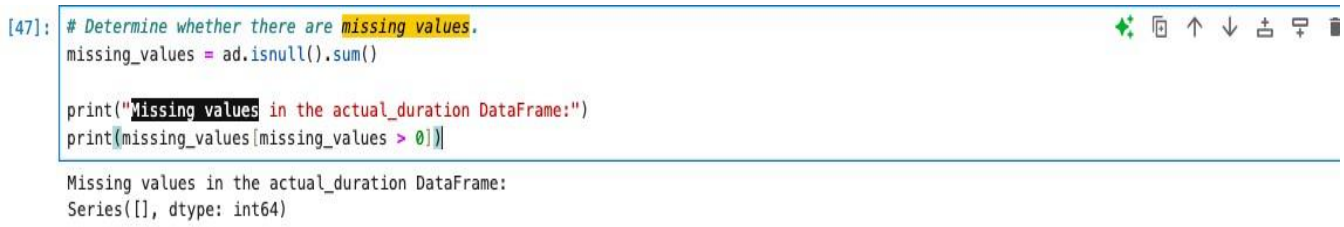
```
# Change the date format of ad['appointment_date'].
ad['appointment_date'] = pd.to_datetime(ad['appointment_date'])


# View the DateFrame.
ad

# View the head
ad.head()
```

	sub_icb_location_code	sub_icb_location_ons_code	sub_icb_location_name	icb_ons_code	region_ons_code	appointment_date	actual_duration	count_o
0	00L	E38000130	NHS North East and North Cumbria ICB - 00L	E54000050	E40000012	2021-12-01	31-60 Minutes	
1	00L	E38000130	NHS North East and North Cumbria ICB - 00L	E54000050	E40000012	2021-12-01	21-30 Minutes	
2	00L	E38000130	NHS North East and North Cumbria ICB - 00L	E54000050	E40000012	2021-12-01	6-10 Minutes	
3	00L	E38000130	NHS North East and North Cumbria ICB - 00L	E54000050	E40000012	2021-12-01	Unknown / Data Quality	
4	00L	E38000130	NHS North East and North Cumbria ICB - 00L	E54000050	E40000012	2021-12-01	16-20 Minutes	

2.2) Missing Values:

Missing Values were handled by using code in the 'actual_duration' Data Frame. Figure 2 shows the code used to identify the missing values. The 'ad.isnull().sum' part of the code checks for the missing values in each column. 'isnull()' gives a Data frame with the same shape as 'ad' which all the cells are 'True,' however if the corresponding value in 'ad; is missing then it will come up 'False' The 'sum()' will then add up all the 'True' values (>1) which will then result in a series where the index is the column names and values are the count of missing values in the column.

(Figure 2)



2.3) Filtering

Data was filtered to focus the analysis on specific services. This was used when needing to see the amount of GP visits. This allowed for easy analysis as it filtered out any unnecessary data that did not need to be used at the time. This code was used to filter the data for GP visits, 'nc_filtered = nc[nc['service_setting'] != 'GP'].'

2.4 Grouping Data

Data was grouped using 'appointment_month,' 'Service_setting' and 'count_of _appointments. Doing this allowed for summary of the number of appointments and identifying trends over time without needing to do several separate codes. Figure 3 below shows a Data Frame contains the total number of appointments for each month, sorted in descending order by the number of appointments. Therefore, the code helped to quickly identify the busiest months for total number of appointments.

(Figure 3)

```
[568]: # Number of appointments per month == sum of count_of_appointments by month.
       # Use the groupby() and sort_values() functions.
       nc_group = nc.groupby('appointment_month')[['count_of_appointments']].sum().sort_values('count_of_appointments', ascending=False)
       nc_group.head()
```

[568]:

appointment_month	count_of_appointments
2021-11	30405070
2021-10	30303834
2022-03	29595038
2021-09	28522501
2022-05	27495508

2.5 Visualisation

Various graphs were employed to visually represent the data. The primary charts utilized for this project were line charts since they are useful for showing trends in data, such appointment counts over time. Since line charts are arguably the most useful for examining trends over time (Slutsky, 2014), they were the primary chart used.

3) Visualisation and Insights

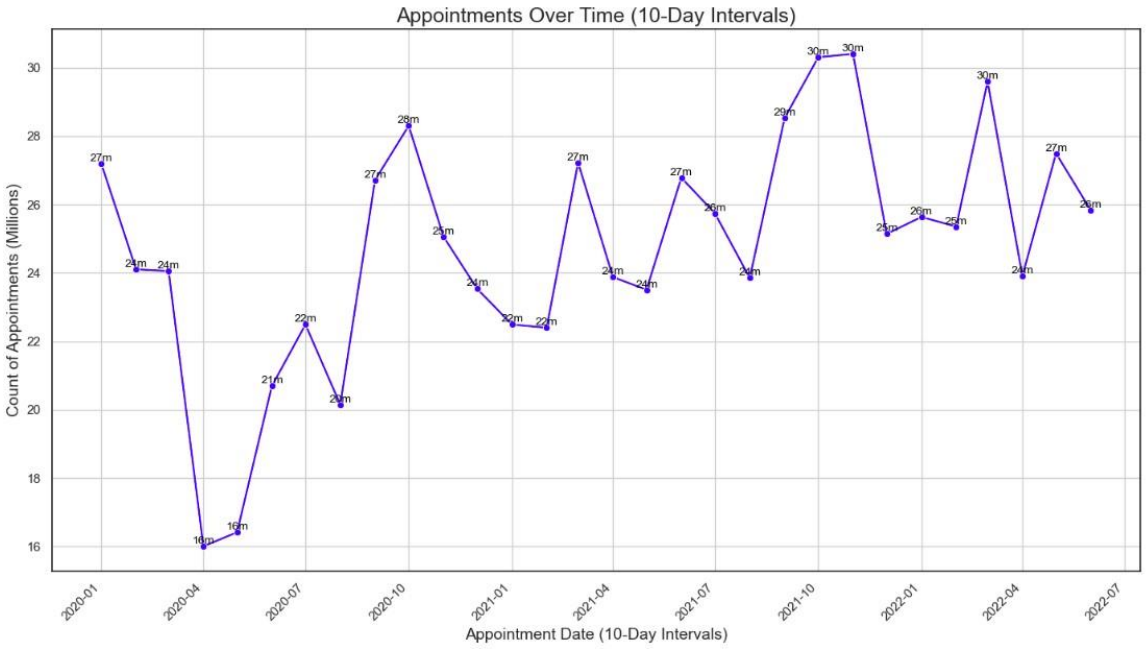
The Visualisations created were aimed at solving the appointment management issue in the NHS. This includes reducing missed appointments, effectively using recourses, and helping patients efficiently use the NHS services.

3.1) Figure 4 below shows a line graph of appointments over time in ten-day intervals. This graph was created to show how appointment volumes fluctuate within each month from 2020. The data was broken down into ten-day intervals so that the stakeholders could identify the short-term peaks, which may link with external factors such as public holidays.

The line chart illustrated that certain intervals within each month peaked. For example, consistently the first ten days of the month especially during January, March and September peak which shows a trend where patients may prefer to schedule appointments at the start of their month, so they do not have to worry about it later.

Additionally, the last ten days of December have high volumes (30million appointments) which could be due to the end year and patients not wanting to carry on their unwanted illnesses onto the next year.

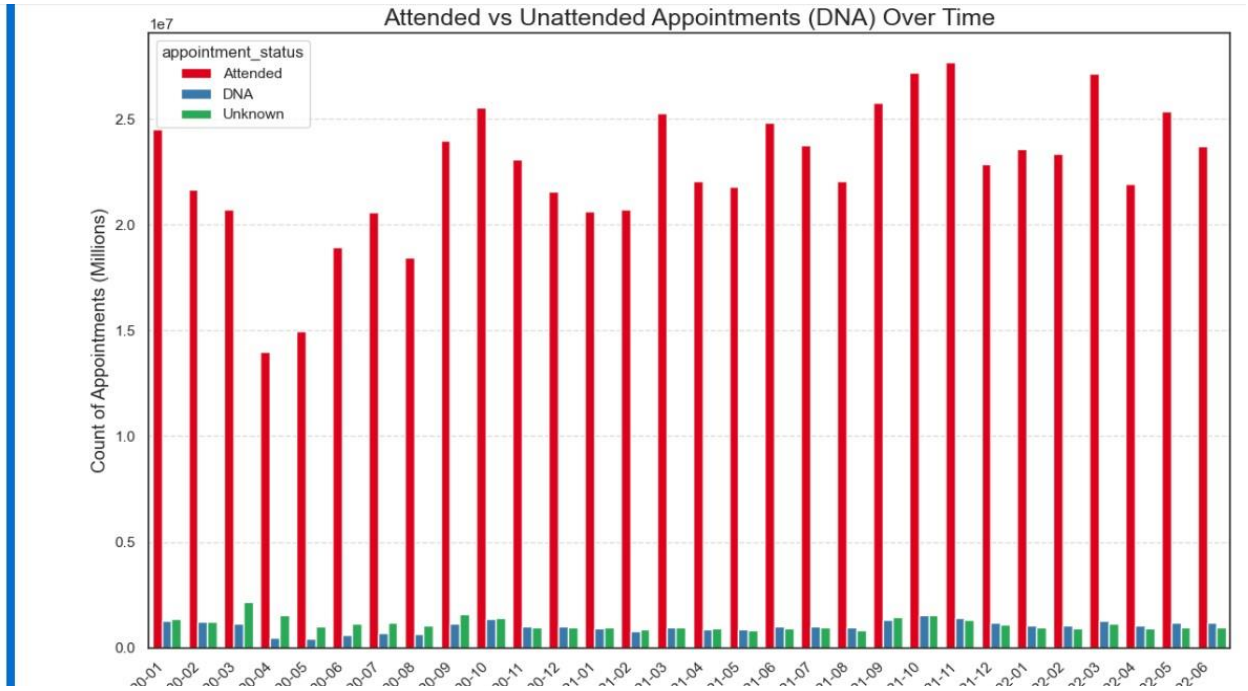
(Figure 4)



3.2) Figure 5 below illustrates attended vs unattended appointments over time. It shows some crucial trends which will impact the NHS' recourses and patient care. In July and August there is a 20% increase in unattended appointments which could be due to summer holidays. On the other hand, January shows a huge peak in attended appointments, this shows that these months do not need as much attention as the 'summer holiday' months.

Moreover, these insights align with the NHS' business objectives of reducing the number of missed appointments and ensuring recourses are allocated properly. This will give stakeholders an indication of how to manage patient flow and reduce the number of missed appointments.

(Figure 5)



4) <u>Patterns and Predictions</u>

The analysis has shown various key trends and patterns in relation to the NHS' operational efficiency. One noticeable trend is the increase in DNA appointments during the summer months- this may be due to the holiday season. This suggests that the NHS needs to target these months with sending their patients constant reminders. Therefore, inevitably reducing the number of missed appointments.

However, January and March consistently show high appointment attendance which could be due to patients worrying about their health after holidays. This trend illustrates that these months may need increased staff as this is when patient appointments are in demand.

The difference in appointment volumes throughout service contexts is another important finding: emergency services and outpatient clinics have significantly larger volumes than other settings. This demonstrates the necessity of resource optimisation to better meet demand in these busy areas, such as staff reallocation or hourly increases.

Following the results from the analysis the NHS should investigate the specific reasons for the high DNA rates during certain periods whilst also investigating the patients' demographics to better understand their attendance to appointments. Furthermore, they should analyse their patient reminder services to reduce the number of missed appointments by possibly implementing a digital engagement tool for patients.

Bibliography

1. U.S. Economic Development Administration (EDA), 2016. *History of the EDA*. Available at: https://www.eda.gov/archives/2016/50/history/#:~:text=In%201965%2C%20Congress%20passed%20the,succeed%20the%20previous%20ARA%20organization. [Accessed 6 August 2024].
2. U.S. National Library of Medicine, 2014. *Electronic health records: an overview*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4078179/ [Accessed 6 August 2024].

Appendix

Figure 1

```
# Change the date format of ad['appointment_date'].
ad['appointment_date'] = pd.to_datetime(ad['appointment_date'])


# View the DateFrame.
ad

# View the head
ad.head()
```

	sub_icb_location_code	sub_icb_location_ons_code	sub_icb_location_name	icb_ons_code	region_ons_code	appointment_date	actual_duration	count_o
0	00L	E38000130	NHS North East and North Cumbria ICB - 00L	E54000050	E40000012	2021-12-01	31-60 Minutes	
1	00L	E38000130	NHS North East and North Cumbria ICB - 00L	E54000050	E40000012	2021-12-01	21-30 Minutes	
2	00L	E38000130	NHS North East and North Cumbria ICB - 00L	E54000050	E40000012	2021-12-01	6-10 Minutes	
3	00L	E38000130	NHS North East and North Cumbria ICB - 00L	E54000050	E40000012	2021-12-01	Unknown / Data Quality	
4	00L	E38000130	NHS North East and North Cumbria ICB - 00L	E54000050	E40000012	2021-12-01	16-20 Minutes	

Figure 2

```
[47]: # Determine whether there are missing values.
      missing_values = ad.isnull().sum()

      print("Missing values in the actual_duration DataFrame:")
      print(missing_values[missing_values > 0])

Missing values in the actual_duration DataFrame:
Series([], dtype: int64)
```

Figure 3

```
[568]: # Number of appointments per month == sum of count_of_appointments by month.
       # Use the groupby() and sort_values() functions.
       nc_group = nc.groupby('appointment_month')[['count_of_appointments']].sum().sort_values('count_of_appointments', ascending=False)
       nc_group.head()
```

[568]:

appointment_month	count_of_appointments
2021-11	30405070
2021-10	30303834
2022-03	29595038
2021-09	28522501
2022-05	27495508

Figure 4

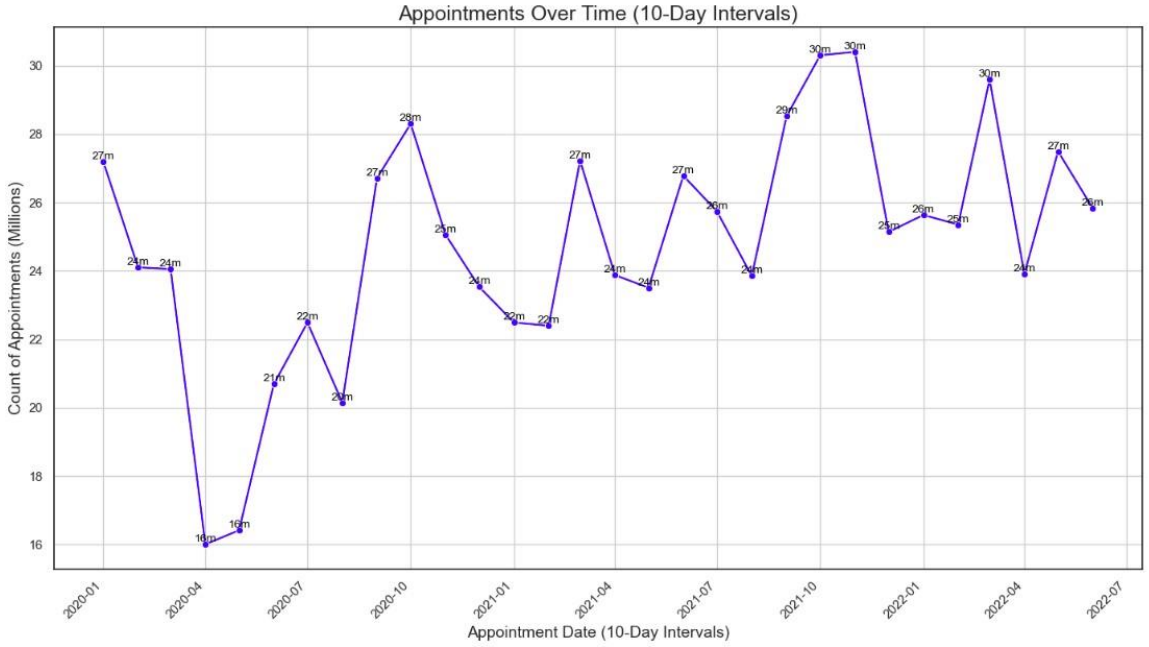
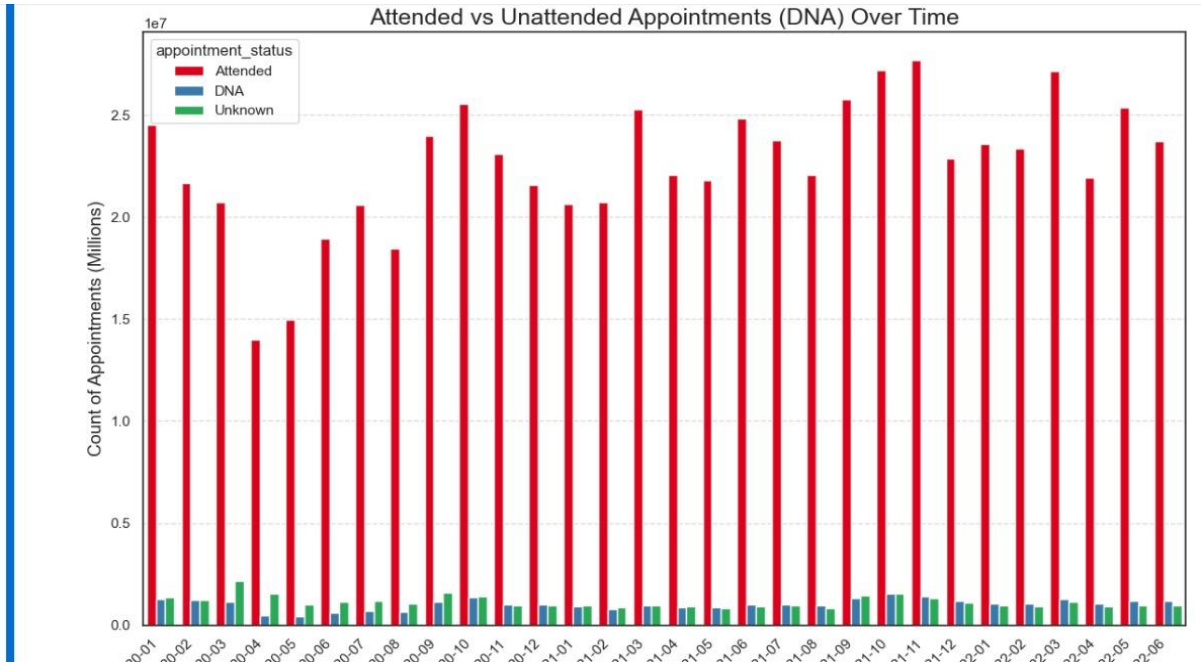


Figure 5


Attended vs Unattended Appointments (DNA) Over Time
1e7
appointment_status
Attended
DNA
Unknown
Count of Appointments (Millions)
2.5
2.0
1.5
1.0
0.5
0.0