BACKROUND

The NHS a National Health Service publicly funded in England has been facing a lot of patients missing their appointments, and the reason for this is yet to be better understood. Missed appointments can lead to worse care for patients, inefficient use of staff, increased waiting times and financial impact. The NHS wants to have a better understanding of the reason for missed appointments and the following questions answered, has there been adequate staff and capacity in the networks? And What was the actual utilisation of resources? As part of the Data analyst team contracted by the NHS, I am going to help in analysing and finding if there are any trends or similarities of patterns for missed appointments.

Analytical approach

Importing Packages

Importing data allows the upload of a dataset from external sources and further enables to perform actionable insights on the dataset to solve or answer business problems. Before performing any actionable insight on the dataset such as cleaning, wrangling and visualising data. The first thing that was done was to import the required packages/libraries into Jupiter notebook first as shown below.

```python
import pandas as pd
import seaborn as sb
```

The reason for choosing pandas was because it has a lot of useful methods that are well suited for manipulating data and seaborn was used for data visualization purposes.

Data Cleaning

The first part of this step was to import the dataset. This was done using the pandas library as shown below.

```python
#load csv national categories csv file
df =pd.read_csv("national_categories.csv")
df['appointment_date'].dtype
```

A similar code snippet was used in loading the other three datasets. After loading the data, the next step was to check if the data contains invalid values like nulls or missing values and do clean ups if necessary. An example of that is shown below.

```python
#print number of missing values in ad
ad.isna().sum().sum()
```

```
0
```

Data Analysis

The first step of this phase is to check the datatypes of the columns of the dataset, stats and the metadata of the dataset. An example of that is shown below.

```
[11]: #print data type in nc
      nc = pd.DataFrame(nc)
      print (nc.dtypes)
```

```
appointment_date        datetime64[ns]
icb_ons_code                    object
sub_icb_location_name           object
service_setting                 object
context_type                    object
national_category               object
count_of_appointments            int64
appointment_month               object
dtype: object
```

The next step was to get a sample of the data as shown below. This helps to give a good idea of what the dataset is about.

```
In [18]: import pandas as pd

         nc = pd.read_excel("national_categories.xlsx")
         nc
```

Out[18]:

|	appointment_date	icb_ons_code	sub_icb_location_name	service_setting	context_type	national_category	count_of_appointments	appointment_month
0	2021-08-02	E54000050	NHS North East and North Cumbria ICB - 00L	Primary Care Network	Care Related Encounter	Patient contact during Care Home Round	3	2021-08
1	2021-08-02	E54000050	NHS North East and North Cumbria ICB - 00L	Other	Care Related Encounter	Planned Clinics	7	2021-08
2	2021-08-02	E54000050	NHS North East and North Cumbria ICB - 00L	General Practice	Care Related Encounter	Home Visit	79	2021-08
3	2021-08-02	E54000050	NHS North East and North Cumbria ICB - 00L	General Practice	Care Related Encounter	General Consultation Acute	725	2021-08
4	2021-08-02	E54000050	NHS North East and North Cumbria ICB - 00L	General Practice	Care Related Encounter	Structured Medication Review	2	2021-08
...	...	...	...	...	...	...	...	...
817389	2022-06-30	E54000054	NHS West Yorkshire ICB - X2C4Y	Extended Access Provision	Care Related Encounter	Unplanned Clinical Activity	12	2022-06
817390	2022-06-30	E54000054	NHS West Yorkshire ICB - X2C4Y	Extended Access Provision	Care Related Encounter	Planned Clinics	4	2022-06
817391	2022-06-30	E54000054	NHS West Yorkshire ICB - X2C4Y	Extended Access Provision	Care Related Encounter	Planned Clinical Procedure	92	2022-06
817392	2022-06-30	E54000054	NHS West Yorkshire ICB - X2C4Y	Extended Access Provision	Care Related Encounter	General Consultation Routine	4	2022-06
817393	2022-06-30	E54000054	NHS West Yorkshire ICB - X2C4Y	Extended Access Provision	Care Related Encounter	General Consultation Acute	19	2022-06

817394 rows × 8 columns

Data visualization

The final phase of this section was the data visualization part. This was useful in revealing some general insights about the dataset like the sample shown below, the visualization revealed the peaks months for appointments and utilisation based on count of appointments respectively.

```
sns.lineplot(x = "appointment_month", y = "count_of_appointments", data=nc_ss)
plt.show()
```
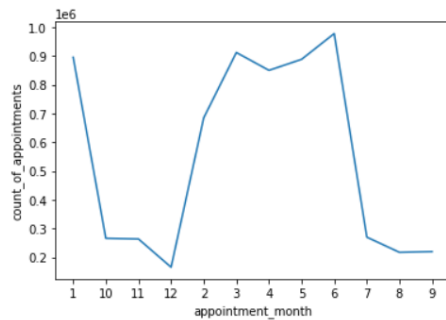


```
In [17]: #visualisation for number of visits based on appoinment type
         sb.lineplot(x=df['appointment_mode'],y=df['count_of_appointments'])
Out[17]: <AxesSubplot:xlabel='appointment_mode', ylabel='count_of_appointments'>
```
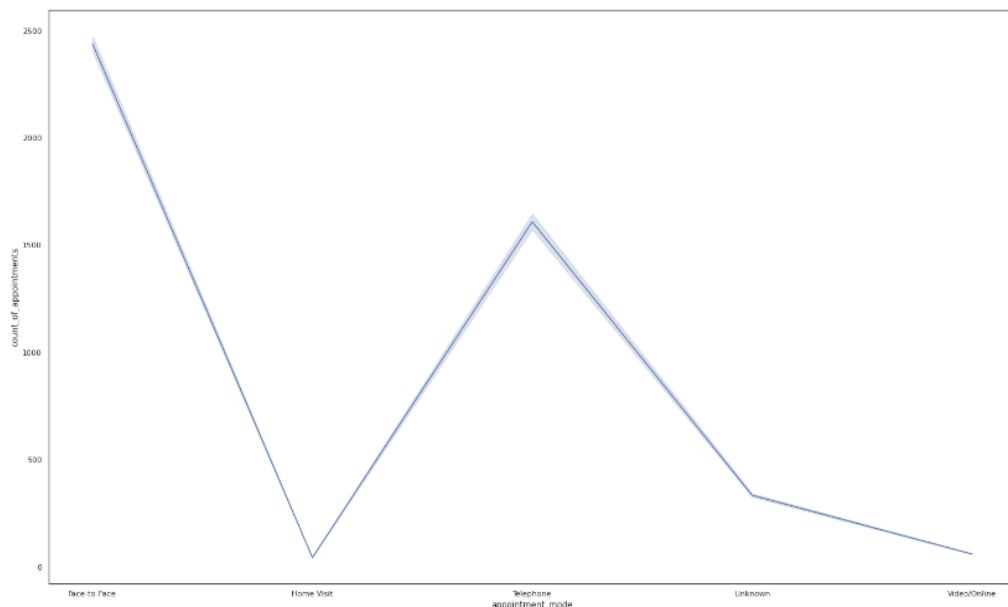

Visualisation and insights

The visualisations used in this data analysis translate findings into a visual context, such as graphs and bar charts, to make data easier for our target audience to understand and pull insights from. The goal of the visualisations is to make it easier to identify patterns, and trends in the data set. The following charts were used as visualisations, line and bar charts. First, a line chart was selected to compare the appointment modes count of appointments; from one of the visualisations, the highest mode of appointment is face-to-face. If investigated further, it will likely entail a higher probability of people missing their appointments as it has a higher count of appointments. To name a few, this could be due to a few reasons such as someone being in isolation or having difficulty with travelling.

The time of the season, such as spring to summer, can also affect appointment attendees. In addition, another line graph was used to display the count of appointments by month. From this visualisation, the highest count of appointments was found in March, with a total count of 27,170002. Around this time, many people tend to feel sick due to season change sickness, which makes a lot of people book appointments they would not attend after feeling better before their appointment date. In addition, another line chart was used to analyse the highest count of utilisation per month. It can be seen that this occurred between October 2021 and November 2021. This could be due to being in lockdown and many patients suffering from long-term covid symptoms. The higher the number of appointments booked, the higher the number of unattended.

A bar chart was also used to compare the quantitative data from the service setting. The highest type of service setting is GP meaning many patients tend to have this form of appointment. This helps direct the focus on the most used form of appointment and investigate how many people attend their appointments. Good data visualisation should communicate a data set clearly and effectively using graphics. These visualisations make it easy to comprehend the service settings, counts of appointments and service utilisation data at a glance. Complex information has been broken down in a way that makes it simple for the target audience to understand and on which to base their decisions.

Patterns and predictions

From the data analysed it can be seen that GP face to face appointments had the highest count of appointments. The highest count of appointments was found in March and the highest utilisation of service of found between October 2021 to November 2021. From the line graph visualisations it can be seen that face to face appointments were high, this is because before a patients is given a face to face appointment a phone appointment is placed to evaluate if the patient requires a GP face to face appointment, the higher the probability in the telephone appointments means there is going to be a higher probability in the face to face appointments, this suggests there is a clear correlation between the two mode of appointments. Predictions are GP appointment will always be highest form of mode of appointment and the highest counts of appointments will be around spring and summer and the highest utilisation will be around October to November. Recommendation will be the NHS Should do further analysis on staffing levels using appropriate data and a longitudinal study should be carried out across the UK to be able to detect areas of development and changes in patient demographics will help in better understanding the reason for missed appointments.