Emmanuel\_ElizabethDA201\_Assignment\_Notebook.ipynb

GitHub is very useful as an experimental tool while testing and verifying ideas. It is useful in hosting project and explore findings, maintain and keeping knowledge bank of codes for the project which can be more useful in future. It can also assist in keeping your work safe from accidental overwriting.

GitHub is a very popular way of building sharing and showcase your work with colleagues and potential employees as part of a portfolio. GitHub is a place to explore and manage your work and as a portal to your portfolio. In summary, GitHub is an ideal place for building your portfolio, control changes over time, storing your code, and learning from and with others.

The data wrangling enables to answer the business question and to reshape the data, more specifically the shape of the dataset, the description of the numerical variables. The data visualisation was done observed to understand the relationship between the variables and potential outliers. This also includes ensuring there is no white space, ensured the consistency of the letter casing and checked for typographical errors. The data visualisation stats well before we create a graphic or image. During the data exploration it was observed that most of the content in the cell exceeded the lines and therefore needed to expand the width to accommodate the contents by selecting the column and expand. Align cell A (Appointment date) to the left through the alignment tab at the top.

Align cell G (Count of appointment) the appointment cell was all align to the left and was sorted into descending order through the data tab, sort and filter group, then largest number to the smallest

The same process was used for to realign both the Appointment region and Actual duration, using this exploration there are missing values, column name re adjustment, the data was cleaned and then analysed before imported to the Jupiter notebook.

The presented problem identified that NHS incurs significant cost when patient miss general practitioner (GP) appointments. The reason needs to be better understood by investigating why the patients do not attend. While reducing or eliminating missed appointment would be beneficially as well as socially, the government needs a data informed approach to decide how best to handle this problem. Has there been adequate staff and capacity on the networks. The main question is, what was the actual utilisation of resources?

In solving the above problem, groupby was mostly used for grouping and aggregating data, this was used to summarise data, the groupby function allows you to effectively split data by a specified categories or group, the category ranges from locations, Appointment month, Count\_or\_appointment.

Groupby was used to solve the problem by groupby the appointment month, find and calculate all the appointment which has January and then sum up the total.

Groupby was also used to calculate the month column to sum up the value for January, sum up all the appointments and do the same for January and March.

The appointment month was grouped together and then added the count of appointment so that we can arrive at a value.

The value is then divided by the new variable of 30. This was done by writing the data frame name and column name and then divide by 30. The result is then used to calculate the utilization of appointment out of 1,200,000 appointments that NHS can handle. Communicate that story to the audience

Every effective visualisation tells a story, the visualisation communicates that story to the audience and enable the audience to engage at a deeper level, it can help people to understand the data implications and it can make the data to be more relevant and memorable.

However, Matplotlib is a steppingstone towards building great visualisations using python.

Figures and axes are the starting point for plotting any chart or graph using Matplotlib in python.

A figure gives a holistic view of the chart, which includes the smallest elements of a chart such as title, legends, plot, subplots, axes and so on.

Axes are a part of figures which define the subplot of a chart. It is the actual plot that you are drawing the region that renders the data set

Starting the visualisation of this assignment, the Matplotlib library was imported at the beginning and ran. The variables was also ran and the data was then put to plot

Matplotlib.pyplot was used to write pythonic code. The subplots () function of matplotlib (plt ) as plt.subplots()

The observation is that while ploting on a simple graph, the value of s axis and y axis is of importance to avoid the python assigning default value. In the assignment, it was also observed that the same formular can be used for the similar ones, in most cases there appear to be syntax errors and one reason was because of letter casing, another trending error is (name not define).

In presenting the analyses of the case scenario, I gathered that the data was inadequate because the General Practitioner (GP work between Monday -Friday while the other service settings don’t work.

The General Practitioners (GP) do face to face appointments while other settings do not operate likewise. The General Practitioners (GP) do the same appointments where other settings do not. The accuracy of the dataset and the information presented in a coherent and cohesive manner can play a major role in arriving at credible and efficient information that is needed to inform intelligent decision making.