Assignment: Diagnostic Analysis using Python

By Timothy Ayling

Background/context of the business:

Contracted by the NHS to explore the reduction of missed appointments through a data informed approach.

NHS questions:

- Adequate staff and capacity in the networks?
- What was the actual utilisation of resources?

Investigative analysis:

1. What is the number of locations, service settings, context types, national categories, and appointment statuses in the data sets?
2. What is the date range of the provided data sets, and which service settings reported the most appointments for a specific period?
3. What is the number of appointments and records per month?
4. What monthly and seasonal trends are evident, based on the number of appointments for service settings, context types, and national categories?
5. What are the top trending hashtags (#) on Twitter related to healthcare in the UK?
6. Were there adequate staff and capacity in the networks?
7. What was the actual utilisation of resources?
8. What possible recommendations does the data provide for the NHS?

Analytical Approach:

I opened GitHub and opened a new repository selecting the public setting. I saved my Jupiter files (py and ipynb) in the attachments and continued to do so whilst working on my assignment. I eventually had three files saved there namely my report pdf file, my Jupyter Notebook ipynb file and my presentation recording in mp4. The 'Public option' in GitHub allows for a shared function available for team members and others to work from and use the data. I also gave further description to the file for easier selection and search parameters.

I imported the csv and xlsx extension files into Jupyter through the upload function. I then wrote code in the program to read it in a data frame. I imported pandas into the program Shape dtypes columns head tail missing values.

I determined the number of locations, service settings, context types, national categories and appointment statuses in the data sets.

I determined the number of appointments and records per month. I then determined the monthly and seasonal trends based on the figures I achieved from the above outcomes.

Twitter: Max and min value and I sorted the data in ascending order.

Good day – I would like to apologise as I really wanted to finish this course. I have been unwell for many weeks somewhere around 6 weeks. I am currently sick again whilst writing this unfortunately. I have been off work in bed mainly. I have been on an oxygen chamber 4 x and other medication which didn't work.

Symptoms included extreme fatigue, fever, coughing, vertigo

Is there any way of me having this referred for another date.

Regards

Timothy Ayling

Describe the approach taken to import, clean, and analyse data in Python. Include a detailed and insightful description of the processes you used and the decisions you made during analysis, such as the choice of libraries, functions, and variables. Ensure that the description of the steps taken to prepare data for analysis is clear, well organised, and relevant to the given scenario.

Visualisations and insights: 350

Describe the rationale for the selected visualisations, and ensure that the interpretations of the visualisation outputs are detailed, insightful, and relevant to the business objectives.

Patterns and predictions: 200

Clearly articulate any patterns, trends, or insights you discovered. Ensure they relate to the business scenario. Include recommendations on any areas for further exploration

1Prepare your GitHub repository.
You'll set up a GitHub repository to store and update your project files. You will set your repository to **Public** for the duration of the course.
For the final:

- **report and Jupyter Notebook** you will demonstrate your GitHub setup consisting of the load and push updates for all the files. (**Hint:** You can also include screenshots where your GitHub name and file structure are clearly visible.)
- **presentation** you will provide reasons for the possible value and impact of GitHub in the context of the scenario. (**Hint:** The scenario suggests you are part of a team. How might the team benefit from a tool like GitHub?)

1. **Data ingestion, wrangling, and predictive modelling:** Utilise an accurate set of functions, variables, parameters, and attributes to import the data files, wrangle the data, analyse, and visualise using Python.
    -
        - When importing, make sure you import the correct files in the appropriate format.
        - When wrangling, make sure to transform your data into a format that comprises consistent values, uses appropriate data types, and has no missing values.
        - When analysing data, ensure that you utilise the correct Python libraries and give your function and variables names that are intuitive and descriptive.
        - Ensure that you provide detailed and insightful descriptions of code and outputs at each stage in your Jupyter Notebook (using the text markdown feature).

2Import and explore the data.
You'll create a new Jupyter Notebook, import the data files, assess the quality of the data, and conduct some initial analysis to familiarise yourself with the data.
For the final:

- **Jupyter Notebook** you should indicate your process of importing the three data sets with the Pandas library, and storing the data in DataFrames. You'll also need to determine whether there are missing values, column names, data types, metadata, and descriptive statistics.

- **report** you will provide a summary of the decisions you made during the initial exploration.
- **presentation** you will describe the process you followed, indicate what you observed, and explain how you approached the analysis.

3Merge and analyse the data.
You'll continue to work in the Jupyter Notebook you created in Week 2 and perform data wrangling and manipulation to search for answers to general questions that will help you better understand the data.
For the final:

- **Jupyter Notebook** you should document your process of data wrangling and data manipulation. This can be based on the approach and guidance provided.
- **report** you should explain how you approached the general questions, what you found through this part of the analysis and how these initial stages of the analysis contributed to answering the questions posed by the NHS.
- **presentation** you will describe the process you followed, indicate what you observed, and explain how it informed your next steps.

4**Identify and visualise initial trends.**
You'll continue to work in the Jupyter Notebook from Week 3 and create visualisations to identify possible monthly and seasonal trends for service settings, context types, and national categories.
For the final:

- **Jupyter Notebook** you should include a selection of visualisations you created in the search to identify trends as well as your interpretation of the visualisations and rationale for selecting the visualisations you did.
- **report** you need to explain your process and rationale behind the visualisations as well as your interpretation of the outputs.
- **presentation** you will describe the process you followed, indicate what you observed, and explain the trends you've identified based on the visual outputs.

- **Visualisation and predictions:** Ensure that the created visualisations display your findings and communicate trends related to the scenario.
    - When plotting charts, follow the basic visual design principles concerning chart type, colour, size, resolution, and layout.
    - When plotting charts, include extra context and information around the visualisations, such as why you chose the particular visualisation to convey the result and how the audience should interpret the output.

- Provide an explanation of the code you used to prepare the visualisations as well as the interpretation of the outputs at each stage in your Jupyter Notebook.

5**Analyse the Twitter data.**

You'll continue to work in the Jupyter Notebook from Week 4 and analyse Twitter data to identify the top trending hashtags (#) related to healthcare in the UK.
For the final:

- **Jupyter Notebook** you will demonstrate your ability to parse complex data, identify hashtags (#) and keywords, display the contents of the messages, and create visualisations to identify possible trends.
- **report** you will provide a clear and concise explanation of the process you followed in your Jupyter Notebook and the possible value of Twitter data in this scenario.
- **presentation** you will discuss the value of adding external data (background information) to the analysis. This will also be an opportunity to discuss the ethical and data quality considerations related to using data from social media.

6**Make recommendations.**

You'll continue to work in the Jupyter Notebook from Week 5 and perform further data analysis to answer the two questions posed by the NHS and provide your recommendations.
For the final:

- **Jupyter Notebook** you should indicate your process of answering the questions by importing data, exploring data sets, analysing data, and creating suitable visualisations.
- **report** you will clearly articulate a response to the questions posed by the NHS and summarise your analysis by providing recommendations for the NHS. You can also share any obstacles you faced in the process and how you overcame them.
- **presentation** you will describe the process you followed, indicate what you observed, and explain your answers to the questions and possible recommendations for the NHS. You can also suggest further questions to explore.