Data Analytics Career Accelerator Course 2 Assignment: Diagnostic Analysis using Python

Student name: Rose Ryan

Submission date: 17 August 2023 (with extension)

Word count = 959

Declaration:

The work is my own and has been created with academic integrity.

Table of Contents

1. Background/ context of the business 2
2. Analytical approach 3
3. Visualisation and insights 4
4. Patterns and predictions 5

1. Background/ context of the business

The government needs a data-informed approach to decide how to handle the problem of missed appointments, which incurs the NHS significant costs.

The reasons for missed appointments need to be better understood in order to reduce or eliminate them as much as possible.

At this stage of the project the two main questions posed by the NHS are:

- Has there been adequate staff and capacity in the networks?
- What was the actual utilisation of resources?

The purpose of this analysis is for the NHS to establish whether their staffing levels are correct across multiple trusts and across healthcare professional (HCP) type, for a period between 2021 and 2022.

This will help inform whether staffing levels need to be raised or lowered at certain periods for future years, in order to maximize staffing provision.

2. Analytical approach

- The metadata text document file provided for the three data files (*metadata_nhs*) was opened in a text editor and reviewed, which contained details of the data set, data quality, and reference.
- A new Jupyter notebook was created for the project using the Anaconda environment, so Python could be used for the code.
- The GitHub repository which was going to be to be used for the project was added at the top of the notebook in a markdown cell.
- The Numpy and Pandas libraries were imported into the new Jupyter notebook.

- The three semi-wrangled data sets provided (*actual_duration.csv, appointments_regional.csv* and *national_categories.xlsx*) were imported into the Jupyter notebook.
- The data sets were given unique and appropriate aliases (*ad*, *ar* and *nc*) upon importing, which are short in length to enable easy and fast code writing.
- Three new DataFrames (*ad, ar* and *nc*) were created to store the three data sets.
- The three new DataFrames (*ad*, *ar* and *nc*) were sense-checked to determine the metadata, including column names, number of rows and columns, data types, and whether there were any missing values.
- The descriptive statistics and metadata of the three DataFrames were determined using the *df.describe())* and *df.info())* functions.

- The data sets were explored to include: number of locations; top five locations by number of records, number of service settings, context types, national categories, and appointment statuses.
- Functions used to explore the data included the *value.counts(), count(),* and *print()* methods.

- The date ranges of the *ad* and *nc* DataFrames were established.
- The date formats of the DataFrames were changed to datestring to enable easier analysis of the data.
- The most popular service setting for NHS North West London from 1 January to 1 June 2022 was determined.
- The month with the highest number of appointments was determined.
- The total number of records per month was determined.

- The correct Python libraries were utilised, and functions and variables given names that are intuitive and descriptive.
- Detailed and insightful descriptions of code and outputs at each stage were provided using the text markdown feature at each stage.
- PEP 8 was adhered to.

3. Visualisation and insights

- The Seaborn and Matplotlib libraries were imported to enable visualization of the DataFrames.
- Visualisations were created in order to identify monthly and seasonal trends in the data.
- When plotting charts, the basic visual design principles concerning chart type, colour, size, resolution, and layout were followed.

- The data type of the appointment month was changed to string to allow for easier plotting.
- Records were aggregated on a monthly level to determine the sum of records per month.

- The appointments were plotted over the available date range in order to review the service settings for months, using the Seaborn lineplot function.
- The appointments were plotted over the available date range in order to review the context types per month, using the Seaborn lineplot function.

- Four visualisations were created to indicate the number of appointments for service setting per season. The seasons are summer (August 2021), autumn (October 2021), winter (January 2022), and spring (April 2022).
- All seasonal graphs showed a drop over weekends, apart from the graph for autumn (October 2021) where there appear to have been appointments in the *Planned Clinical Procedures* category over the weekends. This could be looked into further.

- There was adequate staffing over the time period looked at, with capacity reaching a maximum of ~80%.

- Analysis of Twitter data, and specifically hashtags during the time period being looked at, did not yield any correlation with appointment or attendance trends. As the Twitter accounts were not connected to the patients, this data unlikely to yield any helpful trends.

4. Patterns and predictions

- The output for appointments plotted to review the service settings for months showed that the *General Practice* and *Unmapped* categories had the highest number of records overall.
- For every context type, there was a rise into Autumn, then a steep drop from November to December (which could be attributed to Christmas holiday period), then they remained relatively steady until the end of the period being looked at, in June.

- All seasonal graphs showed a drop over weekends, apart from the *autumn* time period, where there appear to have been Planned Clinical Procedures over the weekends.

- There was adequate staffing over the time period looked at, with capacity reaching a maximum of ~80%. Staffing levels could be revisited and scaled down as this implies there would have been staff present and being paid when they did not need to be, or that staff could be reallocated to other functions or roles.