A Jupyter Notebook (as an IPYNB file) in which you demonstrate your code and note your observations and justifications using Markdown. Use the following naming convention for your file: LastName\_FirstName\_DA201\_Assignment\_Notebook.ipynb.

Senario:

You are part of a team of data analysts that was contracted by the National Health Services (NHS), a publicly funded healthcare system in England. The team has been provided with internal and external data and a number of high-level business questions concerning the utilisation of services, missed appointments, and the potential value of using external data sources such as Twitter (now rebranded as X) . Your role is to refine the business questions to actionable analytic questions based on your review of the available data. You will then plan and perform the exploratory data analysis and present the insights to the respective stakeholders.

The NHS must expand its infrastructure and resources to match its increasing population capacity. For this, it needs to budget correctly. When deciding on budget allotment, the NHS must understand the utilisation trends of each component in its network.

Some stakeholders feel that the NHS’s capacity should be increased while others feel that, based on current trends in utilisation, the current capacity is adequate and that efforts to make better use of existing infrastructure and resources are sufficient. Your team has been tasked to explore the available data and offer suggestions and recommendations based on your observations.

Therefore, reducing or eliminating missed appointments would be beneficial financially as well as socially. The government needs a data-informed approach to decide how best to handle this problem. 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?

As you are new to Python, your role on the team is to get started with data exploration, data wrangling, creating visualisations, and identifying possible trends in the data sets. For this, you’ll investigate several questions:

What is the number of locations, service settings, context types, national categories, and appointment statuses in the data sets?

What is the date range of the provided data sets, and which service settings reported the most appointments for a specific period?

What is the number of appointments and records per month?

What monthly and seasonal trends are evident, based on the number of appointments for service settings, context types, and national categories?

What are the top trending hashtags (#) on the supplied Twitter data set and how can this be used in the decision-making process?

Was there adequate staff and capacity in the networks?

What was the actual utilisation of resources?

What insights can be gained by looking at missed appointments?

What are the most important patterns visible in the data relating to the use case?

What insights can be gained from the data, and what recommendations can be made to the NHS based on these insights?

Important

Each assignment activity is based on the knowledge you acquire during the module. The assignment activities are prompts intended to serve as a guide to help you complete your final assignment. You may decide to omit some of the work you complete for the activities. You may also explore beyond what is suggested in these activities. The decision regarding what to include and exclude should be based on actionable insights to support the business case. To complete the assignment, you’ll draw upon your Python, and data analytics skills to deliver a technical report and Jupyter Notebook to technical stakeholders

Completing your analysis

|  |  |
| --- | --- |
| **Module** | **Assignment activity** |
| **1** | **Exploring the data and planning your analytic approach**  You’ll set up a dedicated folder on your own computer and familiarise yourself with the provided data files, metadata, and the Jupyter Notebook template. You will re-review this main assignment activity page and document your anticipated approach to analysing the data.  The following should be present in your submission at the end of the course:   * **Jupyter Notebook:**Demonstrate good practice in terms of adding code comments and observations as markdown. * **Report:**Demonstrate sound logic in your analytic approach, a clear articulation of the business questions, and a clear overview of the important decisions and findings. * **Presentation:**Summarise the business context and break down business questions into actionable elements that can be used to align the business audience. (**Hint:** In many cases your presentation will be attended or reviewed by stakeholders not familiar with the project, or who may have different expectations. It is therefore important to provide a succinct version of the context and business question to help align stakeholders and manage expectations.) |
| **2** | **Importing and exploring the data**  You’ll update the provided template or 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.  The following should be present in your submission at the end of the course:   * **Jupyter Notebook:** Demonstrate 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. Code should contain comments stating the purpose of the code and observations should be documented in Markdown cells. * **Report:** Provide a summary of the decisions you made during the initial exploration that are relevant and important for technical users that want to validate and reproduce the results presented to business users. All insights presented to the business should be based on observations from the data that are described in the technical report. (**Hint:**The technical report should contain a brief description, as well as supporting figures and tables, to help technical users navigate the code in the Jupyter Notebook, essentially a roadmap and justification of the important components in your Notebook.) * **Presentation:** Present insights and findings based on basic data exploration of the available data sets as well as a brief description of the analytic process in terms relevant to business users. |
| **3** | **Analysing the data**  You’ll continue to work in the Jupyter Notebook you created in Module 2 and perform data wrangling and manipulation to search for answers to general questions (both supplied and the ones you identified in Module 1) that will help you better understand the data. Note that Module 3 is concerned with data preparation while Module 4 continues the exploratory analysis using visual techniques.  The following should be present in your submission at the end of the course:   * **Jupyter Notebook:**Document your process of data wrangling and data manipulation. This can be based on the approach and guidance provided. Code should be commented and observations should be included in Markdown cells. * **Report:** Explain how you approached the general questions, comment on the data quality and usefulness of the various data sources as well as preliminary findings during the initial stages of the analysis. * **Presentation:** Describe basic observations of the data to provide context for the analysis. Examples could include trends observed in appointments across service settings, context types, national categories, as well as descriptive statistics that may be useful in decision-making processes. |
| **4** | **Visualising and identifying initial trends**  You’ll continue to work in the Jupyter Notebook from Module 3 and create visualisations to identify possible monthly and seasonal trends for service settings, context types, and national categories.  The following should be present in your submission at the end of the course:   * **Jupyter Notebook:** Include a selection of visualisations you created during your exploratory data analysis and those you intend to present to the business. The code should contain comments to state the purpose of the visualisations and observations and interpretations should be included as Markdown cells..(**Hint:**After exploring the data you may need to refine the visualisations for specific target audiences both in terms of the types of visualisations selected and data or perspective. An example may be outliers being relevant to technical users but not necessarily to business users.) * **Report:** Explain your process and rationale behind the visualisations as well as your interpretation of the outputs.You should also explain any refinements made to visualisations to ensure that your message can be clearly communicated to business stakeholders in the presentation. * **Presentation:** Describe the observations and insights relating to the use case in business terms using appropriate visualisations that can be used to support your observations and suggestions. |
| **5** | **Analysing NHS-related Twitter data**  You’ll continue to work in the Jupyter Notebook from Module 4 and analyse the provided Twitter sample data to demonstrate the potential usefulness of data from external sources such as social media.  The following should be present in your submission at the end of the course:   * **Jupyter Notebook:** 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:** 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. You should also make recommendations regarding how data collection and analytic efforts could be improved. * **Presentation:** 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** | **Making recommendations**  You’ll continue to work in the Jupyter Notebook from Module 5 and perform further data analysis to answer the two questions posed by the NHS and provide your recommendations.  The following should be present in your submission at the end of the course:   * **Jupyter Notebook:** Indicate your process of answering the questions by importing data, exploring data sets, analysing data, and creating suitable visualisations. * **Report:** 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, how you overcame them, and list your recommendations for further improvements. * **Presentation:** Describe observations and insights using visualisations (data-driven storytelling) to explain your answers to the questions and possible recommendations for the NHS. You can also suggest further questions to explore based on your analysis of the data. |

Ensure that your code and outputs within your Jupyter Notebook leverage best practices related to data ingestion, wrangling, visual design principles, making predictions, and reproducibility of code.

Note that you will submit your Jupyter Notebook and while this is a working document, technical users will refer to this to determine the standard and accuracy of code.

The suggested approach for your analysis is as follows:

1. **Data ingestion, wrangling, and predictive modelling:**Utilise an accurate set of functions, variables, parameters, and attributes to import the data files, and then wrangle, analyse, and visualise the data 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).
2. **Visualisation and predictions:**Ensure that the created visualisations display your findings and communicate trends related to the scenario appropriate to the intended purpose and audience (exploration and technical audience versus business stakeholders).
   * 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, who the intended target audience is, 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.
3. **Jupyter Notebook:**Organise a clean, easy-to-understand, and structured Notebook that has a logical flow that communicates your end-to-end analysis.
   * Ensure your code and outputs have corresponding descriptions so that the reader understands what each of the components in the Jupyter Notebook means, or is meant to do.
   * Make sure the code lines in the Notebook have sequential numbering so that the readers can interpret the workflow and logic. (**Hint**: Restart the kernel, run the code, and review the outputs before moving on to submitting your work.)
4. **Ensure** that you use the correct naming convention before you submit. Save and submit your files as:
   * LastName\_FirstName\_DA201\_Assignment\_Notebook.ipynb