Assignment Activity 1: Exploring the data and planning your analytic approach

This assignment activity will help you to successfully complete the final assignment. The provided instructions will guide you to the minimum expected activities to complete during your exploratory data analysis and the presentation of your insights. You are encouraged to think critically and apply logic to identify patterns and trends that the business can use  to answer specific questions. You will also use these insights to inform both the technical report and business presentation you’ll submit at the end of the course.

**Scenario**

**Recall the work you’re doing for the NHS for your final assignment. Review the requirements introduced in**[**Assignment: Diagnostic Analysis using Python**](https://fourthrev.instructure.com/courses/895/assignments/2971)**. As part of your work, you need to set up your working environment on your local computer and plan your approach.**

**Objective**

Set up your environment and plan your analytic approach to organise everything you need for the project.

**Approach**

1. Download the LSE\_DA201\_Assignment\_files.zip file from [Assignment: Diagnostic analysis using Python](https://fourthrev.instructure.com/courses/895/assignments/2971).
2. Unzip the file to access the data.
3. Explore the contents of the zipped file and document your approach and observations:
   1. Open the metadata\_nhs.txt file with a text editor (e.g. Notepad), and familiarise yourself with the metadata of the three CSV files.
   2. Open the provided assignment notebook template and review the content at a high level.
4. Plan your approach and document your thoughts (as explored in 1.1.5 Approaching the analysis). You can use a separate document or use comments/Markdown directly in the provided assignment template):
   1. Review the scenario.
   2. Summarise the questions in the provided scenario.
   3. Plan your initial approach for the different questions. (What data do you need and how will you go about answering the question?)
   4. Note that the provided template will guide you in terms of the typical sequence and components expected to be present. You can complete the relevant sections on a weekly basis and use the template as your working document throughout Course 2.
5. Back up your work to a safe location. This would allow you to revert to a previous state in the case of making a mistake in the code, or deleting a section by mistake. (A simple way of doing this is to save or mail a compressed version at frequent intervals).

**Important**

It is important to note that in most cases you will have to spend time formulating and refining business questions rather than simply accepting the scenario as accurate, complete, and practical. As an analyst, it is your task to translate business questions into questions that are suitable for analytical tasks and to provide suggestions regarding a practical approach on how to solve the business need and navigate the opportunities and pitfalls identified. Use the following questions as a guide:

* Can the question be answered?
* Do you need to break the question down into smaller actionable elements?
* Which data source (and columns) are required for the various questions?
* Do you need to clean the data?
* Do you need to subset or merge data sets?
* Are there other risks that need to be raised with the planned approach?
* Are there other opportunities in the data that may be of interest to the organisation that are related to the use case?

This is typically an iterative process that will be refined as you interact with the data and learn more about the use case.

**Criteria:**

**Analytic approach (10%): Compelling insights into the approach to analysing the data are provided. The rationale behind the chosen analysis techniques, the limitations of the data, as well as the chosen techniques and approach, are also included. Business questions are broken down into actionable elements and there is a clear description of the data requirements and approach used to answer the individual questions.**

**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.)**