Note: This assignment contributes 25% to your final SIT112 mark. It must be completed individually and submitted to **Resources and Assessment** section on Moodle before the due date/time mentioned above.

The theme for this assignment, and the subsequent data science project, is to explore data related to Australia. In particular, we will use data provided by the Government at http://data.gov.au. Our data strategy and task specifications for this assignment will focus on the analysis of Medicare office locations in March 2018. Please go to the website of the dataset for more information https://data.gov.au/dataset/location-of-medicare-offices

1. Data and Resources

In the Assignment 1.zip file, you will find the following files:

Filename	Description
medicare-office-locations.csv	This is the dataset file provided by data.gov.au.
field_description.pdf	This file contains description for attributes in the data file.
datadictionary_template.xlsx	This is the template for the data dictionary file in Excel.
assignment1_notebook.ipynb	This is the Jupyter notebook which has been prepared and pre-filled for you to complete the Programming task.

These are the files you will be required to work with for this assignment.

2. Task Description

There are two main tasks for this assignment:

- Construction of the data dictionary (**35 marks**) and
- Programming tasks to perform basic data analysis (**65 marks**).

2.1 Construction of the Data Dictionary (35 marks)

For a data scientist, after obtaining the dataset, the first most crucial task is to obtain a good understanding of the data he or she is dealing with. This includes examining the data attributes (or equivalently, data fields), seeing what they look like, what is the data type for each field, and from this information, determining suitable analysis tools. *A systematic approach to this process, as we have learned from the lectures and practical sessions, is to construct a data dictionary for the dataset.*

Your task is to construct a data dictionary for the dataset you are working with (Medicare office location dataset) using the provided data dictionary template.

You are required to prepare two sheets in your data dictionary Excel file:

- Dataset description [**5 marks**]
- Attribute dictionary [**30 marks**]

The total marks for this task is 35 marks. The data description sheet is worth five (5) marks. The attribute dictionary is worth 30 marks where each correct attribute specification is worth 2.5 mark. Name your solution as *[YourID]_datadictionary.xls* and submit this file.

2.2 Programming task (65 marks)

A Python notebook file **assignment1_notebook.ipynb** has been prepared for you to complete this task. Download this notebook, load it up and follow instructions inside the notebook to complete the task.

The total mark for this task is 65 marks. *You are required to submit your solution in two formats, 1) Jupyter Notebook format and 2) its exported version in HTML.*

3. Summary for submission

This assignment is to be completed individually and submitted to the corresponding Moodle Assignment1 submission link by the due date. **Your submission must be made as a compressed file named [Your ID]_Assignment1.zip** that includes the following files, named the given format:

1. **[YourID]_datadictionary.xls**: your solution for the data dictionary for the business establishment dataset.
2. **[YourID]_assignment1_solution.ipynb**: your Jupyter notebook solution source file.
3. **[YourID]_assingment1_output.html:** the output of your Jupyter notebook solution in HTML.

For example, if your student ID is 123456, you will then need to submit following three (3) files:

- **123456_datadictionary.xls**
- **123456_assignment1_solution.ipynb**
- **123456_assignment1_output.html**

END OF ASSIGNMENT DESCRIPTION