# Data Visualisation and Presentation Assessment (MSCDSA07)

The summative assessment for this module is a written report which accounts for 100 % of the marks. The deadline for the submission is **4 May 2021**, 13:00 oclock and the submission word limit is 4000 words.

## 1. Tasks

### 1.1. Data Selection

You are required to choose a dataset and carry out an exploratory visual analysis to better understand the shape and structure of the dataset.

Three useful links for choosing a dataset is provided under the assessment section.

You need to share your dataset in OneDrive by selecting "**Anyone with the link can edit"** option to customize the link options (Share OneDrive files and folders).

### 1.2. Exploratory Analysis

Perform an exploratory analysis of the dataset using visualisation tools to develop preliminary insights then formulate at least **three questions research** questions.

The research questions should be relevant to the dataset and seek to gain an overview of the shape and structure of the dataset for example:

- What variables does the dataset contain?
- How are they distributed?
- Are there any notable data quality issues?
- Are there any surprising relationships among the variables?

For each question, start by creating a visualisation that might provide a useful answer. Then refine the visualisation (by adding additional variables, changing sorting or axis scales, filtering data, etc.) to develop better perspectives and explore unexpected observations.

You should repeat this process for each of your questions to revise your questions or branch off to explore new questions if the dataset permits.

### 1.3 Visualisations

Visualisations should be created from the dataset, accompanied by a descriptive caption, describing the insight(s) and the story.

Provide justifications for the choice of visualisations, key insights gained during analysis and enough detail for each caption such that anyone could read through your report and understand what you've learned. You are free, but not required, to annotate your visualisations to draw attention to specific features of the data. Your report must consist of captioned visualisations and should convey insights gained during your analysis.

## 2. Deliverables

1. Report named with your student number including the link of your dataset in OneDrive (1234567.pdf)

2. Python source code named with your student number (1234567.py)

**Notes:**
**1:** The assessor(s) will run your code so the code should run without issues on the university machines and produces the repored output. You will not be assessed on the quality of coding.

**2:** You are required to submit both parts for a valid submission otherwise you will fail the assessment.
**3:** If you don't include the link for your dataset you will fail the assessment

**4:** If your code doesn't run and produces errors on the university machines, you will fail the assessment.

**5:** Make sure your code is independent of dataset location. Marker(s) will copy your dataset from OneDerive to the same folder as your Python code.

## 3. Report Structure

The report should be structured according to the followings:

Table of Contents

1. Introduction
2. Dataset
3. Exploratory Analysis
4. Body Sections
5. Findings and Conclusions

Appendices

A: Copy of source code as text (not screenshots)

B: Link of your dataset in OneDrive

References

**Note:** *All the sections/subsections must be numbered except, Appendices and References. All the figures/tables in the report must have captions (e.g. Figure 1, Table 1), properly referenced and explained in the report.*

**Assessment Criteria**

| | Data Visualisations (60%) | Exploratory Analysis (30%) | Presentation and Structure (10%) |
|---|---|---|---|
| 80%+ | The data visualisations have been chosen appropriately, clearly and concisely annotated with a clear justification of the choices in the explanations including colour etc.<br><br>The story within the dataset is presented accurately and concisely in a manner that is widely accessible.<br><br>Python code runs without errors and produces all the reported results. | Focused, concise, analytical reflection on the decision-making process evidencing excellent understanding<br><br>Makes appropriate, insightful, and powerful connections between the questions and relevant theory and empirical data. | 1. Excellent structure<br><br>2. Excellent presentation<br><br>3. Excellent writing skills<br><br>4. Precise, full and appropriate references. |
| 70%+ | For the most part, the data visualisations have been chosen appropriately with correct annotations but justification in the report is not clear.<br><br>The story within the dataset is accessible.<br><br>Python code runs without errors and produces all the reported results. | Focused, concise, analytical reflection on the decision-making process evidencing good understanding.<br><br>Identifies and outlines connections between the questions and relevant theoretical and empirical research from the dataset and other sources; the connections identified are clarified. | 1. Very good structure<br><br>2. Very good presentation<br><br>3. Very good writing skills<br><br>4. Very good and appropriate references. |
| 60%+ | The data visualisations are mostly correct explanations but there are omissions and explanations not always correct.<br><br>Python code runs without errors and produces most of the reported results. | Analytical reflection on the decision-making process evidencing good understanding<br><br>Strong connections are made between the questions and the dataset. | All (4):<br><br>1) Presentation is good but basic<br><br>2) no typos<br><br>3) good writing skills |

| | | | |
|---|---|---|---|
| | | | 4) referencing complies with the Harvard style |
| 50%+ | Inappropriate choice of some of the visualisations however overall, the story within the dataset is understandable, but explanations are incorrect.<br><br>Python code runs without errors and produces some of the reported results. | Limited reflection on the decision-making process evidencing some understanding<br><br>Connections are made between some of the questions and the dataset. | 3 out of 4 |
| 40%+ | An attempt has been made to produce visualisations, but the results are incomplete and confusing with little explanations.<br><br>Python code runs without errors and produces some of the reported results. | Limited reflection on the decision-making process evidencing some understanding.<br><br>The connections made between questions and the dataset are weak or inaccurate. | 2 out of 4 |
| 30%+ | The visualisations are confusing and incomplete with inappropriate choice of colour and layout.<br><br>Python code runs with errors and the reported results cannot be produced. | Poor or no reflection on the decision-making process.<br><br>Makes no connection between the questions and the dataset. | 1 out 4 |