

## 7CS034 Assessment

### Data Science

## ASSESSMENT OVERVIEW

Your assessment consists of two parts each contributing 50% of the module grade. You should upload both pieces of work to the correct portals on canvas before the deadline.

(A) Coursework (prepared using Quarto).

(B) Presentation with Online Submission (prepared using PowerPoint).

Both pieces of work must be based on the dataset provided.

## THE DATA

You will find the dataset for use in this assessment on canvas. The data consists of a modified snapshot of data collected during a household census conducted in England in 2021.

- **ID** - Household identification number.
- **Person\_ID** - Individual ID number for persons in the household.
- **Age** - Age.
- **Mar\_Stat** - Marital Status.
- **INC** - Annual income in pounds.
- **Female** - A binary variable which signals whether or not the individual has identified as female (0 = yes).
- **H8** - A binary variable signaling the response to question H8 on the census “Are all the rooms in this accommodation, including the kitchen, bathroom and toilet, behind a door that only this household can use?”; 1=no, 0=yes.
- **Eth** - Ethnicity.
- **Highest Ed** - Highest level of study.

## (A) COURSEWORK PREPARED IN QUARTO

You must develop an overview and **ggplot** based visual investigation of the dataset using the tools discussed during this module. I'm being deliberately vague here because I want you to **tell me something interesting** about this dataset and the visual relationship between the variables. You'll be required to upload three files to canvas for this part

- A **.qmd** file in the format **StudentID.qmd**.
- A **.html** file in the format **StudentID.html**.
- A zipped folder containing the images, style files which accompany your **StudentID.html** file.

I'll explain each of these in detail in lectures. The number of lines in your **StudentID.qmd** file **must not exceed 200**. The rubric used to mark the assessment is below.

- (i) Appropriate use of **Quarto** and **ggplot** to produce a visual investigation of the dataset.

**[30 Marks]**

- (ii) Appropriate and informative **graphical visualizations** of the relationship between variables in the dataset and a **commentary** on same. **[40 Marks]**

- (iii) An appropriate and informative narrative throughout the coursework which highlights the novel points of interest obtained in your investigation.

**[30 Marks]**

## (B) PRESENTATION WITH ONLINE SUBMISSION (PREPARED USING POWERPOINT).

For this part you are required to upload a PowerPoint presentation outlining the highlights of your investigations from part (A). This should be no more than **five slides** not including title slide and references. The rubric used to mark this assessment is below.

- (i) Appropriate **title slide and use references**.

**[25 Marks]**

- (ii) Consistent & compelling narrative/story around the **graphical findings** highlighted.

**[50 Marks]**

- (iii) A brief critique on the limitations of the dataset in answering questions linked to the findings you have highlighted.

**[25 Marks]**

## SUBMISSION

You should upload your completed assessments to the portals on canvas before the due date shown on your own canvas. I do not reproduce the due date here because for some of you it will be different due to extensions, extenuating circumstances etc.