CM50266 – Applied Data Science
Semester 2 Coursework
Tesco Grocery 1.0 Dataset vs. Income

Initial Submission (90%)   **Set**: 6/2/2024 (Week 19)   **Due**: 19/4/2024 (Week 29), 8pm   **Peer Review Submission (10%)**   **Set**: 29/4/2024 (Week 31)   **Due**: 3/5/2024 (Week 31), 8pm
Percentage of overall unit mark: 50**%**
Submission Location: Moodle/Panopto   **Submission Components**: Video/Executive Summary/Code   **Submission Format**: The video can be any format that Panopto accepts, the executive summary must be a pdf. The code should be uploaded as a zip file.
Anonymous Marking: Y

1   Overview

In this case study you will analyse a large dataset, the Tesco Grocery 1.0 Dataset[1] along with other datasets. You will use existing tools or software you develop to address a series of tasks. You will deliver your case study in the form of a video (oral) presentation of up to ten minutes and a one-page executive summary.

1.1   Dataset

The dataset records the grocery purchases made in Tesco stores across the Greater London area for a year. A paper providing an overview of the dataset is available on Moodle. This was taken from:
https://www.nature.com/articles/s41597-020-0397-7
The dataset is available here:
https://figshare.com/collections/Tesco_Grocery_1_0/4769354

2   The Tasks

2.1   Task 1: Describe the Dataset

In this task you need to describe the dataset. It contains many fields. A good description will find ways

to group fields together to summarise the content rather than simply list all the fields. You need to explain the value of the data rather than only provide basic description. You should also address any assumptions or limitations that need to be considered which may affect how the data can or should be used.

2.2 Task 2: Visualise the Dataset

In this task you will explore the dataset and use one or more appropriate visualisations to extract **two** useful insights from it. You are free to choose both the topics for exploration. The visualisations should conform with the best practice you studied in Semester 1. You are free to create programs and use any libraries or other tools to assist you in this task. You are also encouraged to make use of the skills you have developed in other units such as statistical methods or machine learning approaches where you consider these appropriate.

2.3 Task 3: Combining Datasets

In this task, you will need to identify a source of income data at a suitable area scale over several years to allow comparison of the two datasets. Begin by providing a brief overview and analyse of the new dataset. This could include identifying any trends or looking for geographic trends.

Then use the two datasets to identify any patterns in buying behaviour that appear related to income. It will help you to focus your exploration if you conduct a brief literature review of the topic. In presenting your conclusion you should relate it to the literature you have used. Note that you are also free to identify where a pattern that might be expected is not present.

2.4 Task 4: Presentation

The work you have carried out will be assessed by means of a video presentation. The content of the presentation will be the focus of the assessment, not the production value. However, the presentation should be both concise and clear. It is sufficient to provide an audio commentary over a slideshow. It must be your own voice. All external resources/literature should be properly cited. The presentation must contain a mix of audio, visual and textual presentation that is appropriate to the content.

2.5 Task 5: Executive summary

You should summarise the key points of your case study presentation in a one-page executive summary. Key points include important results, assumptions, and conclusions. This document should also contain your bibliography. The bibliography will not count towards the one-page limit.

2.6 References

References should be in Bath Harvard format when listed in a bibliography or can be in an abbreviated Bath Harvard format or a precise web reference when included within the video presentation.

3 Submission

You must submit three things by the initial deadline.
1. An .mp4 to Panopto, by following the link on Moodle.
2. A .pdf of the one-page executive summary and bibliography to a Moodle assignment.

3. A zip file containing any code you developed for the task. You should not include any data that you downloaded from elsewhere, or large data sets that can be re-created from the code. This will also be uploaded to a Moodle assignment.

3.1 Submission Deadline

The deadline for video, summary and code upload is 19th April 2024, 8pm. You should note that you will be uploading your work to more than one university system and should allow sufficient time to do so.

4 Criteria for Marking

Assessment will be based on the video (oral) presentation and executive summary. Your goal is to deliver a critical and informative submission. The video must be no longer than 10 minutes. It must be submitted as an mp4 that will be playable on the University PanOpto platform. The executive summary should be no longer than a page (~11pt body text) and be in pdf format.

The video should not exceed 10 minutes as markers will not be expected to include content beyond this in their assessment. You are free to use less time if you do not need it. Similarly, only the first page of the executive summary will be assessed (excluding the bibliography.)

Assessment Criteria		
Question	**Criteria**	**Marks**
Task 1	Has the dataset been concisely summarised?	7
	Have any relevant limitations, assumptions or biases been identified in the data?	5
Task 2	Have two useful insights been extracted from the dataset? (3 marks each.)	6
	Has an appropriate style of visualisation been used for each? (4 marks each.)	8
	Do the visualisations follow best practice in their design? (4 marks each.)	8
Task 3	Has an appropriate income dataset been identified?	5
	Is a brief overview of the dataset been provided?	4
	Has a preliminary analysis of the dataset been provided?	5
	Has additional relevant literature been reviewed?	4
	Has a correlation (or lack of) been identified between the two datasets?	4
	Has a reason for this correlation been presented?	3
	Has this reason been related back to the existing literature?	3
Task 4 (Presentation)	Is the content delivered using an appropriate mix of audio and visual approaches?	4
	If the correct terminology used when discussing the content?	4
	Are all graphs, charts and tables clearly labelled?	4
Task 5 (Summary)	Does the executive summary concisely state the nature of the data being used?	4
	Does the executive summary concisely state the two insights produced?	4
	Does the executive summary concisely state any observation about the relationship of the two datasets and provide a reasonable conclusion?	4

| References | Have appropriate references been provided in the expected style? | 4 |

Figure 1: Marking Criteria

4.1 Restrictions

You are permitted to use any visualisation tools or libraries when creating code you choose for this coursework, so long as they are available under a license that permits you to do so.

You may not use AI based tools, especially those that are generative in nature to create the content for this coursework. However, you are permitted to use such tools to help improve your own work, for example those that assist with spelling or grammar.

Any such tools you use should be noted in a section after any reference section. This section will not form part of the document word count.

5 Feedback

This unit will make use of peer review. This means that after the initial deadline for a piece of coursework you will be allocated the work of three other students to assign a mark. This will allow you to see how others have tackled the same problem. The purpose of this is to expose you to issues you may not have identified for yourself and improve your understanding of the problem being tackled.

You will be provided details of how to download the three submissions. You are expected to examine these and compare them to the assessment specification given in this document. Additional assessment advice will be provided after the deadline to assist with peer review.
You will be given a form where you can submit an entry for each submission you examine. You must submit all the forms by the peer assessment deadline. Your review mark will not be applied to the assessment of others. However, 10% of the unit mark will be determined by how close your reviews match the marks assigned by the lecturers.

You will also receive a review form completed by a lecturer which will provide the initial submission mark.

5.1 Peer Assessment Mark Calculation

The marks your peers give you will not affect your mark for this coursework. Or vice versa. Instead, the 10% of your mark awarded for peer review will be based on how close your assessment is to the mark assigned by the lecturers. To gain these marks your reviews must be complete. A complete review entry means you will have completed a form for each submission allocated to you and provided a justification for each of the criteria you have labelled as not met. The mark awarded will be based on the sum of the absolute difference in marks of each of the three items reviewed, as shown in Figure 2. This means the absolute difference of each of the three marks will be added together not that the difference of three assigned marks will be added together.

Sum of Absolute Differences	Mark Assigned
<= 5	10
<= 10	8

<= 15	6
<= 20	4
<= 25	2
26 +	0

Figure 2: Marks for Peer Review

5.2   Peer Review Deadlines

The work you are to peer review will be released **on 29th April 2024**.

The deadline for submission of the completed review form will be **3rd May 2024, 8pm.**

6   Academic Integrity

Your work will be checked to ensure that you have not plagiarised. For more information about the plagiarism policy at the University see: https://library.bath.ac.uk/referencing/plagiarism

Remember that published work that you refer to in your report should be clearly referenced in your text and listed in a bibliography section given at the end of your report. For more information see, https://library.bath.ac.uk/referencing/new-to-referencing

You should treat downloaded code you have chosen to use within your project in the same way. It should be clear when you have created code yourself, and when the code of others has been included in your work.

7   Extensions

If you require an extension to the deadline, you will need to contact your Director of Studies using the standard form and provide the evidence required given the nature of the extension request. Unit lecturers cannot grant extensions. A gap has been built into the deadlines to allow one-week extensions to be granted without impacting the peer review process.