CI7330 – Data Analytics and Visualisation – Summative Assessment 2021/22

Each student should download their own dataset (CSV file) with their own K number from Canvas. **Be careful to use your own dataset!** They are all slightly different. This system allows you to share ideas and collaborate on the best way to analyse and present your findings. This is itself a valuable learning experience.

You have taken a new job as data analyst at a small supermarket chain. The boss has sent you a dataset and the following email:

> Hi, I have this dataset which was collected for us by the sales team. I want you to use some of your data analytics on it. I don't know much about that sort of thing, so I hope you can explain findings to me in simple terms!
>
> Basically, we have 80 stores in different parts of the UK. I want to see what impact COVID restrictions have had on sales, so we can plan for the future if there is another lockdown. So, the sales people drew out 100 transactions from each store, at random, from a day before COVID happened, and another 100 from a day during lockdown. We have the amount that the customer spent with us in a variable called "spend". The COVID status is in a variable called "covid", and that is either 0 for pre-COVID or 1 for during lockdown.
>
> I got some of the IT people to pull out official government stats on the neighbourhoods where the stores are. The relevant one is called "deprivation". Basically, the higher that is, the more poverty there is in that neighbourhood.
>
> So, what I hope you can tell me is:
> 1. a statistical summary of all the variables
> 2. are they correlated?
> 3. did the spend change between pre-COVID and lockdown? Can you please do one of those tests that check if it is significant? I want to know if the difference is real.
> 4. can we use covid status and deprivation to predict average spend? can you give me a formula that would do that, and tell me how uncertain it is?
> 5. Also, can you give me two graphs, one for question 3, and one for question 4. I want to show them at the next board meeting, so please make them look professional.
>
> Thanks!
>
> Chief Data Officer
> Spendalot Inc.

For this assignment, you should submit one Word document file containing five clearly labelled sections, one for each of the boss's questions. Each section should contain no more than one table of statistics and one paragraph (maximum 120 words), to answer the boss's questions. The fifth section should contain two visualisations, as requested by the boss, and one paragraph (maximum 120 words) that explains them.

The assessment will be marked with 16/26 points for the tables and graphics, and 10/26 for the explanations. So make sure you think carefully about the explanations!

Each student should download their own dataset with their K number from Canvas. Be careful to use your own dataset! They are all slightly different. This system allows you to share ideas and collaborate on the best way to analyse and present your findings. This is itself a valuable learning experience.

But even if you choose exactly the same approach as a fellow student, you will get slightly different results (tables and graphs), so we can see that you are not cheating. Do not simply copy and paste a fellow student's outputs – we will know if this has happened. However, **you must write your own text explanations**.

If you supply more than one table in any section, we will only mark the first one. If you supply more than two visualisations in section 5, we will only mark the first two. And if any of the paragraphs exceed 120 words, we will only mark the first 120.

Good luck!