Data Science (Eng) 774/874
Post-block Assignment 3

April 2023

Kindly complete the following assignment **in groups of 2 OR 3** and submit your assignment as an electronic submission on SUNLearn by **23:55 on 1 May 2023.** You will need to sign up to a group on Sunlearn, specifically for this assignment.

1. Data Science (Eng) 874 AND 774 (postgraduate diploma) students [30]

The final deliverables of this assignment are a data visualization/dashboard and a supporting report in PDF format. You may research and select any data visualization tool and use it to develop your visualization/dashboard. You may use any dataset of your choice for your data visualization/dashboard – datasets provided for other assignments and examples in this course may, however, not be reused for this assignment.

In your report, you need to explain how you used the CRISP-DM methodology to develop your visualization/dashboard. This report should contain detailed information about each phase of the CRISP-DM process from business understanding through to deployment. You are expected to cover all the phases of CRISP-DM (although modeling is excluded for 774 students).

2. ONLY Data Science (Eng) 874 students [10]

This section is only applicable to the MEng module level, Data Science (Eng) 874 (not 774 (postgraduate diploma)), students.

Over and above the requirements described in Section 1, you need to ensure that your visualization/dashboard includes a predictive element, i.e. you should not simply visualize your dataset, but also use machine learning to predict future data points.

3. Evaluation rubric

Your visualization/dashboard will be marked according to the following rubric:

Criteria		**Excellent - above average**			**Not at desired level of competency**
	Weighting	**4**	**3**	**2**	**1**
A clearly defined topic that answers a specific question or facilitates decision making.	5	Visualization has a concise and clearly defined topic that addresses one question.	Topic is well defined but visualization addresses too many questions to be useful.	Topic is somewhat defined and the visualization addresses multiple questions.	Poorly defined topic that addresses too many questions to be useful.
Application of CRISP-DM	15	Accurate, concise description of how each relevant phase of CRISP-DM was executed in the assignment.	Only minor errors in the application of CRISP-DM e.g. data preparation errors.	Major flaws in the application of CRISP-DM e.g. a missing phase.	No or very little reference to the use of CRISP-DM.

Compliance with Tufte's Visualization Aesthetic	5	Outstanding application of Tufte's visualization aesthetic.	Generally competent application of Tufte's visualization aesthetic. Minor errors may exist.	Major violations of Tufte's visualization aesthetic.	No consideration of Tufte's visualization aesthetic.
Level of originality and innovation	5	A truly special and innovative visualization. May allow for user interaction.	Good, creative use of visualization tools.	Limited effort made to select and use data visualization tools.	Run of the mill visualization which could easily have been generated by basic Excel. Not particularly interesting.
ONLY Data Science (Eng) 874 students					
Competent use of prediction algorithm	10	Excellent and highly competent application of a prediction algorithm which is well documented in report.	Generally competent application of prediction algorithms. Minor errors may exist.	Major errors in the use of prediction algorithms.	No or highly flawed prediction of future data points.

4. Example data visualizations

There are many examples of good data visualizations available online. Here is one link that can be used as a starting point:

https://blog.udacity.com/2015/01/15-data-visualizations-will-blow-mind.html