|  |
| --- |
| **Portfolio for Data Science COM618** |
| Your student number and name |
|  |
|  |
|  |

**Note:** This is only an example template to be used, you can use your own as far as you follow the assessment brief.

The reference list at the end does not count towards the word count. Feel free to be creative and support your answers by **adding artefacts such as screenshots, code snippets, diagrams, charts etc., where appropriate throughout the report.** For both part A and B of the assessment, it is important to provide evidences (related work) from the literature throughout the portfolio. Please use mainly journal articles, book, conference proceedings and reports when you do your research and if you need to cite these.

Code snippets/screenshots will act as artefacts, so make sure you add these at every important step you took.

# **Introduction**

* Write about the background of the topic area and a brief literature review, some challenges in the sector and/or the problem statement.

## Aims/Objectives of the coursework- Describe in a few lines the aims and objectives of the work, make reference to a research question if you can.

# **Methods**

Dataset

Your report should justify your choice of datasets and be informed by research. Talk about the dataset, where you got it from (source), what it contains etc. Here you can add a table explaining the original dataset (the one you built in one of the weekly activities).

Give as much information regarding the dataset and about your topic area, work that have been done, detailed use of literature about the topic.

# **Analysis and Results**

Dataset Preparation

Clean the data if and as necessary. Your report should document and justify any techniques you have utilised to assess the quality of the data. Talk about datatypes, some techniques that can be applied etc. You can provide screenshots when necessary. If the dataset did not need any preparation, you still need to mention why; for example, no null values etc, how did you check that.

If you are using a subset of the dataset for preparation (in case you have identified too big datasets), you have to mention this here. A good dataset should contain from a few hundreds to a thousand, you will not be penalized if you are using bigger datasets as far as it does not slow your system during analysis.

Exploratory Data Analysis

Primarily EDA is for seeing what the data can tell us.

Utilise suitable data mining tools and analysis techniques to find significant patterns and trends (SPSS, Excel, Tableau, WEKA, Python libraries, etc.). You do not need to use all the tools but using Python or a mix of tools will provide you with higher grades.

Add histograms, bar chats etc to see if you can get any insights from the data, look at the distribution etc. Look for null values, drop unnecessary columns, rows, find outliers, correlation among variables, clusters etc… This is not limited to what is in this document, the further you deep, the better it is and you should document each step.

Here you can add another table explaining the new dataset if you wish, after making some changes.

Data Modelling and Visualisation

Identify your independent and dependent variables. Perform analysis such as time series, multivariate, bivariate analysis, linear regression etc. You can try to explain what you are trying to predict, follow the code examples from the class activities.

Use appropriate tools to perform some visualisation on the chosen dataset. The choice is yours, based on your future intention of work and also the familiarity of the tool. Your report should document and justify the techniques you have used to mine and analyse the data based on some examples from the weekly class activities as well ad from your own research.

***Evaluation***

After you have performed some modelling and obtained some results, identified any patterns in the data or provided some insights, you can compare your work with existing work from other researchers in a few paragraphs to see what are the effects of one variable on another or among any other variables. Evaluate your work critically explaining what could have done to support your analysis. What approaches could have been used. Look for models/data cleaning/data pre-processing techniques from the literature to compare your approach and evaluate critically.

Limitations and Challenges

# Talk about the challenges. All of you will face challenges regarding your dataset, how did you overcome some of these? What were your limitations?

# **Conclusion**

A few lines to conclude on the work and what you have learnt. The conclusion can be more like a reflective summary. A reflection paper is meant to illustrate your understanding of the material and how it affects your ideas and possible practice in future.

**(<1800<Total word count<2200)**

# **Reference List (Harvard Style)**

***Appendix (if needed)***

# You will need to produce screenshots of data analysis and visualisation results and code snippets where appropriate (your artefacts) within the report. You should also submit your artefacts, i.e, Python codes/ Tableau files/ demo etc in a zipped folder.

# NOTE : You will be expected to demonstrate your artefacts (recorded video up to 7 min long maximum)as specified by the module leader.

You should submit 2 files:

- Your report as a word or pdf document separately uploaded

- Your dataset/source code/Tableau files/ Recorded demo/video/link to video or any data files as a zip file