CHOICE OF A PROBLEM BY TEAM 22

A DATA SCIENCE APPROACH TO FORECAST ELECTRICITY CONSUMPTION IN AUSTRALIA

John Student (z123456), Jim Student2 (zID), Jack Student3 (zID).

School of Mathematics and Statistics
UNSW Sydney

September 2020

Submitted in partial fulfilment of the requirements of the capstone course DATA3001

Abstract

A first assessed activity will be the Choice of a Problem (10%). Each group will have to choose a complex open-ended real-world problem to solve (and related data sets, potentially big, to use) among those presented during the seminar in Week 1. The first step is to choose a topic and delimit the problem which will be studied. You should understand and define the overall problem and propose a solution. Having only a short time to complete the project, it is crucial that the problem that will be studied is well defined. The approach to solve the problem should be original, so it will be necessary to carry out a preliminary literature review. This will prevent any plagiarism. This will also enable you to situate the project in a more global context. At this stage, one can identify potential approaches and software that will be used to solve the problem. It is necessary to plan any simulation carefully and to decide what statistical analyses will have to be carried out. All the sub-steps of the project should be planned precisely (a detailed schedule will be created). A kind of short draft version of your final report, consisting of a two-page proposal (excluding the mandatory pages devoted to the title, abstract, contents and references) will be submitted and discussed with one of the instructors in Week 4 for approval. This implies that this task should be finished a bit before Thursday 4pm in Week 4 to leave your lecturer some time to read it before your discussion. Assessment criteria involve the following: a clear definition of the problem; a clear description of the data format and their storage; a clear description of the relevant data (variables, missing and corrupt values, etc.); the level of difficulty of the chosen data sets (size, complexity, messiness) and its relevance for the chosen problem; the appropriate choice of software and statistical methods to solve your research questions; the precision of your scheduled activities.

Contents

1 Introduction and Motivation

You should describe the problem in your own words and motivate its importance. Your description should be clear for a reader not familiar with the problem (e.g., a student in your discipline but who is not involved in the course, another professor from your school who is not teaching DATA3001). The idea is that you explain clearly what you have understood about what exactly is the problem after having attended the seminar in week 1, and after your discussions with the sponsor and/or your lecturers.

Then, briefly describe how you intend to solve the problem (e.g., compare several existing methodologies, apply specific statistical techniques, etc.).

2 Brief Literature Review

Survey the most related work you have found in the literature. Due to space restrictions, describe only the most relevant work and discuss its connections to your work. Describe the methods employed in the related work, as well as the employed measure of success.

In order to incorporate your own references in this report, we strongly advise you use BibTeX. Your references then need to be recorded in the file `references.bib`, and cited as follows [1].

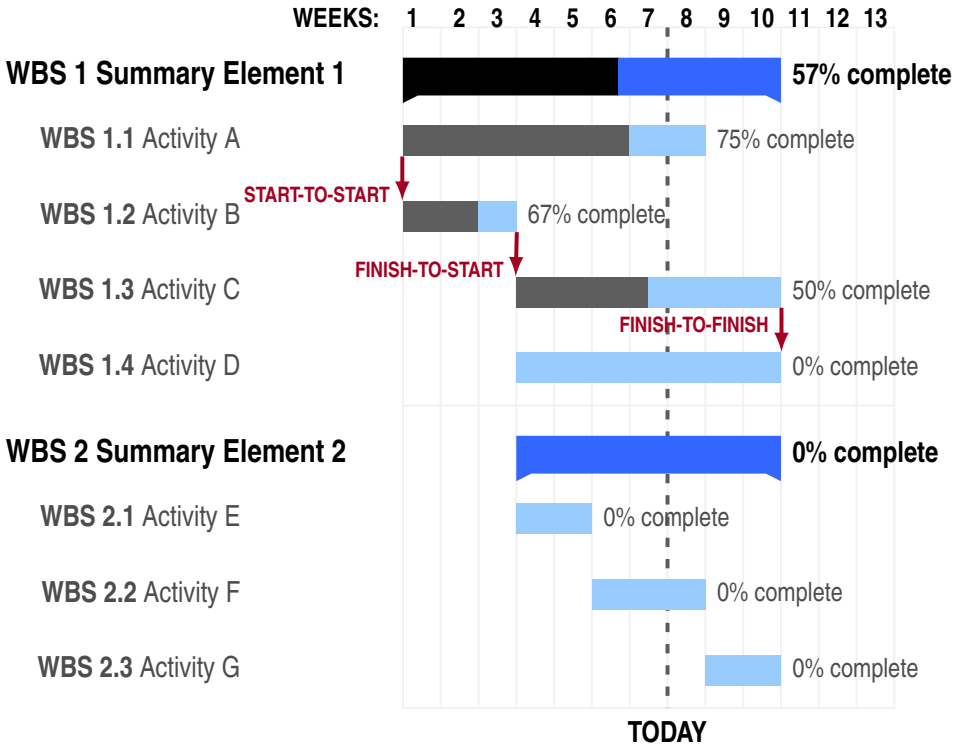
3 Methods, software and Data Description

You should describe the techniques you intend to use and why you have selected these techniques. Also, describe the software and libraries that you will need to implement your analyses. You can revise this later since this is an initial proposal. Finally, provide a clear description of the relevant data (format, storage, variables, messiness, size, complexity) and its relevance for the chosen problem.

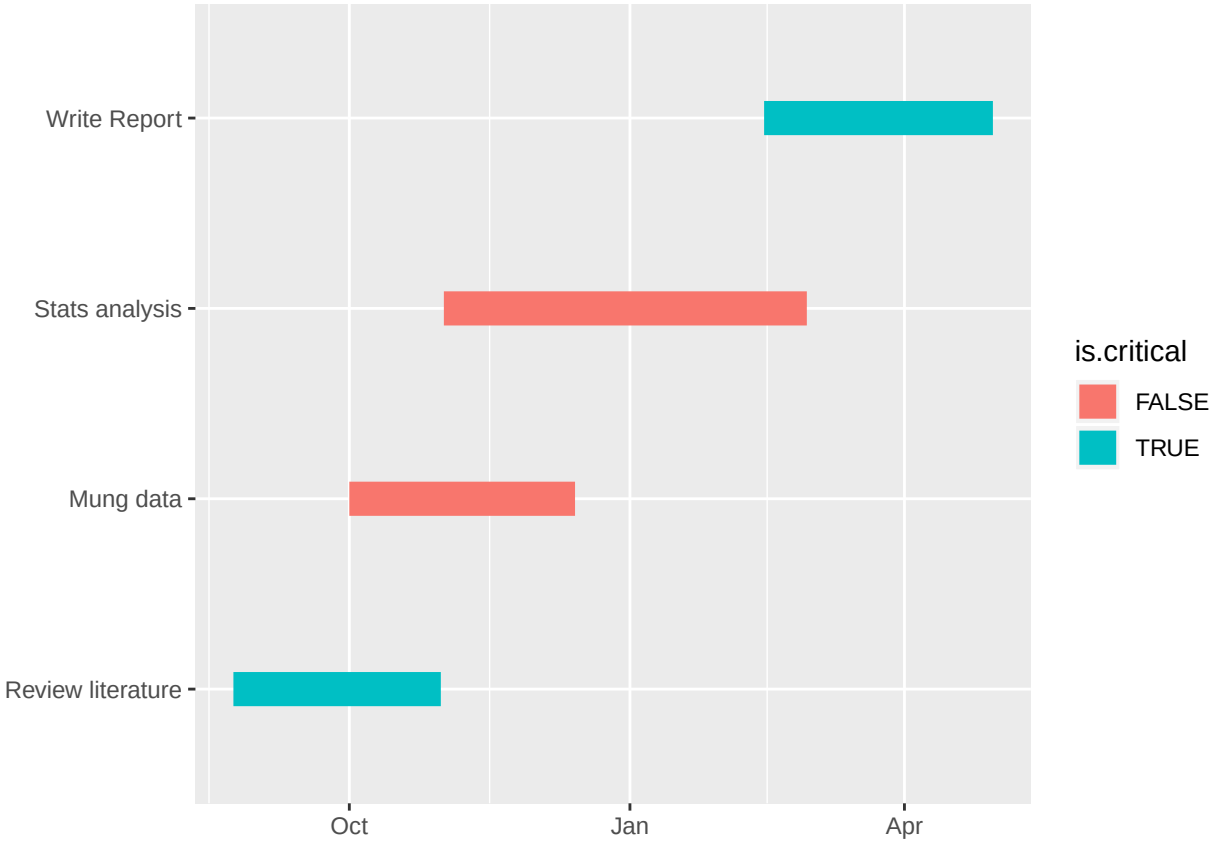
4 Activities and Schedule

Lists the main project activities and create a timetable for the activities. You can use a Gantt Chart (see next page).

Here is an example created using LaTeX:



Here is an example created using R:



References

[1] P. Lafaye de Micheaux, R. Drouilhet, B. Liquet, The R Software: Fundamentals of Programming and Statistical Analysis, Statistics and Computing, Springer New York, 2013.
URL https://books.google.fr/books?id=Ji-8BAAAQBAJ