S&DS 425/625 Report Template

Name(s)

Date

Abstract

An overview of your report, including one or so sentences on each of these:

- a non-technical description of the problem you are trying to solve or the question you are trying to answer, and why you are trying to answer that question
- a non-technical description of the data, where it came from, and what it contains, including possibly the predictors, the outcome, and the observations
- a non-technical description of what kind of analysis you did, including high-level description of what the predictors were, what the outcome was, and how to interpret the results of the model
- a brief summary of the models that are used
- a non-technical description of the results of the model and main takeaways.

An abstract is one paragraph with text only and is aimed at a technical audience. This appears at the beginning of the report.

Executive Summary

An executive summary is typically longer than the abstract, up to a page, could possibly contain key visualizations, tables, or other figures that help communicate either the raw data or the results of the model, and is intended for someone outside of the data science/analytics team of an organization. It is important to be as concise as possible, and describe each of those points above without using language that is overly technical and not part of commonly used English. The executive summary is a separate document.

Note that in the abstract, executive summary, and throughout the report you should avoid using first-person singular pronouns like "I" and "me", even if you are the only author. Use "we" or use passive voice.

Introduction

A few paragraphs that contain the following:

- background on the topic you are studying, including the motivation behind the project and the problem statement you mentioned in the abstract, but in more detail
- a description of the data
- a sentence or paragraph describing what is contained in each section of the rest of the paper. Include roughly one sentence per section, that describes what is in the section and the main takeaways from each section. For example,

"Section 2 contains data exploration and visualization, which reveals that *****. In Section 3, we build several different predictive models and find that *****. We discuss the results of the model, including *****, in Section 4. Finally, we discuss conclusions, recommendations and ideas for future work in Section 5."

There are similarities between the introduction and the abstract. The introduction is longer and more detailed, especially in terms of background, previous work, and motivation of the problem, and contains a brief outline of the contents of the rest of the paper.

Data exploration and visualization

This section will have descriptive statistics and visualizations of the raw data. Use this section to reveal to the reader any interesting relationships in the data, and convince the reader that the predictors are related to the outcome. Visualizations are one of the most powerful ways to communicate information to the reader, so it is important to spend time producing clear, descriptive, eye-catching visualizations.

The package `pubtheme` has a `ggplot` theme called `theme_pub` that helps with making publication-quality visualizations with `ggplot`. See https://github.com/bmacGTPM/pubtheme. There are also several templates there that you can copy, paste, and modify.

If you display a data visualization or some other summary of data, discuss the significance of what you see. What does this tell you about the data? What does it tell you that will help you with modeling? Do not simply show a visualization for the sake of showing a visualization.

Since `echo=F` is the option chosen at the top, the default will be to show the output but not the code:

```
[1] 2
```

If you don't want to show the output either, you can use `include=F`:

Nothing was shown above. If you want to force it to show the code for some reason, you can override the default by putting the options `echo=T` for this chunk.

```
1+1
```

```
[1] 2
```

However, since this is a formal report, you will likely not want to show code.

Modeling/Analysis

Describe regression or classification model(s) used, or the analysis that was performed. For each regression or classification model, discuss

- any assumptions that are made
- the observation, the predictors, and the outcome (aka the rows of X, the columns of X, and y)
- what model you are using, and write out the model
- what the coefficients mean (when applicable) and how this is related to your problem
- appropriate measures of the performance of the model, such as measures of fit and predictive ability
- whether or not you think the model is appropriate for this kind of data, and why, and
- how easy/hard it is to interpret the results and explain them to either a technical or non-technical audience.

For other kinds of analysis, what you give is highly dependent on the type of analysis. But in general, talk about assumptions, if they are appropriate, how they might not be appropriate, and why you chose this type of analysis.

Visualization and interpretation of the results

Create visualizations of the results when appropriate, focusing on visualizations that

- help describe aspects of the results that have real-world interpretation
- help the reader understand how the model addresses the problem you are studying.

Visualizations are one of the most powerful ways to communicate information to the reader, so it is important to spend time producing clear, descriptive, eye-catching visualizations.

Discuss the results of the model or models you chose, and describe how they are related to the problem statement or question that you were trying to answer in the project.

If you have built multiple models or types of analysis, compare the measures of performance and the ease of interpretability across models or types of analysis, stating which model or models performed best, and which model or models were most interpretable. Finally, decide which model or type of analysis is best for your particular problem based on some combination of performance and interpretability.

Conclusions and recommendations

One or two paragraphs stating conclusions, recommendations, and ideas for future work and improvements.

Appendix (optional)

Any supporting information or additional information that isn't necessary to have in the main body of the paper. For example, huge tables can go here, especially if they are more than one page. Tables that are, for example, 100 pages most likely should not be included at all.