



UNSW  
SYDNEY

CAPSTONE PROJECT BY TEAM 22

A DATA SCIENCE APPROACH TO FORECAST  
ELECTRICITY CONSUMPTION IN NSW

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

School of Mathematics and Statistics  
UNSW Sydney

September 2024

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF  
THE CAPSTONE COURSE ZZSC9020

---

## Plagiarism statement

---

I declare that this thesis is my own work, except where acknowledged, and has not been submitted for academic credit elsewhere.

I acknowledge that the assessor of this thesis may, for the purpose of assessing it:

- Reproduce it and provide a copy to another member of the University; and/or,
- Communicate a copy of it to a plagiarism checking service (which may then retain a copy of it on its database for the purpose of future plagiarism checking).

I certify that I have read and understood the University Rules in respect of Student Academic Misconduct, and am aware of any potential plagiarism penalties which may apply.

By signing this declaration I am agreeing to the statements and conditions above.

Signed: \_\_\_\_\_

Date: \_\_\_\_\_

Signed: \_\_\_\_\_

Date: \_\_\_\_\_

Signed: \_\_\_\_\_

Date: \_\_\_\_\_

---

## Acknowledgements

---

TBW...

25/07/2020.

---

## Abstract

---

TBW ...

---

## Contents

---

Chapter 1	Introduction	1
Chapter 2	Literature Review	2
Chapter 3	Material and Methods	3
3.1	Software . . . . .	3
3.2	Description of the Data . . . . .	3
3.3	Pre-processing Steps . . . . .	3
3.4	Data Cleaning . . . . .	3
3.5	Assumptions . . . . .	3
3.6	Modelling Methods . . . . .	3
Chapter 4	Exploratory Data Analysis	4
Chapter 5	Analysis and Results	5
5.1	A First Model . . . . .	5
Chapter 6	Discussion	6
Chapter 7	Conclusion and Further Issues	7
Appendix		9
<b>Codes</b>	. . . . .	9
<b>Tables</b>	. . . . .	9

---

# CHAPTER 1

## Introduction

---

This R Markdown template can be used for the ZZSC9020 course report. You can incorporate R [R Core Team(2017)] chunks and Python chunks that will be run on the fly. You can incorporate  $\LaTeX$  commands.

Before submitting the last version of your report, you might want to use <https://overleaf.com> to collaborate with other members of your team directly on the  $\LaTeX$  version of this document (which is a byproduct you get when you Knit it from studio).

We suggest you organise your report using the following chapters but, depending on your own project, nothing prevents you to have a different organisation.

---

## CHAPTER 2

### Literature Review

---

Here are a few references that can be useful: [Xie et al.(2018)Xie, Allaire and Grolemond] and [Lafaye de Micheaux et al.(2013)Lafaye de Micheaux, Drouilhet and Liquet]. See also <https://bookdown.org/yihui/rmarkdown-cookbook/>

In order to incorporate your own references in this report, we strongly advise you use BibTeX. Your references then needs to be recorded in the file `references.bib`.

---

## CHAPTER 3

### Material and Methods

---

#### 3.1 Software

R and Python of course are great software for Data Science. Sometimes, you might want to use **bash** utilities such as **awk** or **sed**.

Of course, to ensure reproducibility, you should use something like **Git** and RMarkdown (or a Jupyter Notebook). Do **not** use Word!

#### 3.2 Description of the Data

How are the data stored? What are the sizes of the data files? How many files? etc.

#### 3.3 Pre-processing Steps

What did you have to do to transform the data so that they become useable?

#### 3.4 Data Cleaning

How did you deal with missing data? etc.

#### 3.5 Assumptions

What assumptions are you making on the data?

#### 3.6 Modelling Methods



---

## CHAPTER 4

### Exploratory Data Analysis

---

This is where you explore your data using histograms, scatterplots, boxplots, numerical summaries, etc.

```
import numpy as np
np.random.seed(1)
np.random.normal(0.0, 1.0, size=10)

## array([ 1.62434536, -0.61175641, -0.52817175, -1.07296862,  0.86540763,
##        -2.3015387 ,  1.74481176, -0.7612069 ,  0.3190391 , -0.24937038])
```

---

## CHAPTER 5

### Analysis and Results

---

#### 5.1 A First Model

Having a very simple model is always good so that you can benchmark any result you would obtain with a more elaborate model.

For example, one can use the linear regression model

$$Y_i = \beta_0 + \beta_1 x_{1i} + \cdots \beta_p x_{pi} + \epsilon_i, \quad i = 1, \dots, n.$$

where it is assumed that the  $\epsilon_i$ 's are i.i.d.  $N(0, 1)$ .

---

## CHAPTER 6

### Discussion

---

Put the results you got in the previous chapter in perspective with respect to the problem studied.

---

## CHAPTER 7

### Conclusion and Further Issues

---

What are the main conclusions? What are your recommendations for the “client”?  
What further analysis could be done in the future?

A figure:



Figure 7.1: A caption

In the text, see Figure [7.1](#).

---

## References

---

- [Lafaye de Micheaux et al.(2013)Lafaye de Micheaux, Drouilhet and Lique] Lafaye de Micheaux, P., Drouilhet, R., Lique, B., 2013. The R Software: Fundamentals of Programming and Statistical Analysis. Statistics and Computing, Springer New York. URL: <https://books.google.fr/books?id=Ji-8BAAAQBAJ>.
- [R Core Team(2017)] R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- [Xie et al.(2018)Xie, Allaire and Golemund] Xie, Y., Allaire, J., Golemund, G., 2018. R Markdown, The Definitive Guide. Chapman and Hall/CRC. URL: <https://bookdown.org/yihui/rmarkdown/>.

---

## Appendix

---

### Codes

Add you codes here.

### Tables

If you have tables, you can add them here.

Use [https://www.tablesgenerator.com/markdown\\_tables](https://www.tablesgenerator.com/markdown_tables) to crete very simple markdown tables, otherwise use L<sup>A</sup>T<sub>E</sub>X.

Tables	Are	Cool
col 1 is	left-aligned	\$1600
col 2 is	centered	\$12
col 3 is	right-aligned	\$1