Project Report Template

Graham Williams

25th August 2015

1 Introduction

A paragraph or two introducing the project
2015-08-25

2 Business Problem

Describe discussions with client (business experts) and record decisions made and share understanding of the business problem.

3 Data Sources

Identify the data sources and discuss acess with the data owners. Document data sources, integrity, providence, and dates.

4 Datas Preparation

Load the data into R and perform various operations on the the data to shape it for modelling.

5 Data Exploration

We should always understand our data by exploring it in various ways. Include data summaries and various plots that give insights.

```
library(rattle)

## Loading required package:  RGtk2
## Rattle:  A free graphical interface for data mining with R.
## Version 3.5.0 Copyright (c) 2006-2015 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.

library(dplyr)

##
## Attaching package:  'dplyr'
##
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(xtable)
library(dst)
set.seed(42)

dsname  <- "weatherAUS"
ds      <- tbl_df(get(dsname))
nobs    <- nrow(ds)
obs     <- sample(nobs, 5)
vars    <- 2:7
ds      <- ds[obs, vars]
dst <- weatherAUS[sample(nobs, 20), vars]
kable(dst, row.names=FALSE, digits=0, booktabs=TRUE)
```

Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine
Portland	8	14	1	1	0
Woomera	22	39	0	11	12
NorahHead	14	21	0	NA	NA
Townsville	21	30	0	10	12
MountGambier	7	17	0	1	5
MelbourneAirport	7	14	3	1	3
Nuriootpa	12	18	2	9	9
Launceston	5	14	0	NA	NA
WaggaWagga	1	15	0	4	10
MelbourneAirport	12	17	1	9	1
Launceston	12	21	0	NA	NA
Darwin	18	32	0	4	10
Newcastle	NA	20	78	NA	NA
Melbourne	17	30	0	4	11
Dartmoor	0	15	0	1	6
Hobart	4	11	1	0	3
NorahHead	16	24	0	NA	NA
Katherine	15	30	0	8	NA
AliceSprings	16	22	0	5	0
CoffsHarbour	18	25	2	NA	NA

```
print(xtable(ds, digits = 1), include.rownames=FALSE)
```

Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine
Hobart	12.4	21.8	0.2	5.0	9.6
Launceston	9.0	15.3	6.5		
Williamtown	9.3	20.3	0.8	3.6	10.3
PerthAirport	6.9	19.8	0.0	2.2	9.4
GoldCoast	17.7	27.2	1.0		

```
ds <- dst
dst[-1] <- sample(10000:99999, nrow(dst)) * dst[-1]
print(xtable(dst, digits = 0),
      include.rownames=FALSE,
      format.args=list(big.mark=","))
```

Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine
Portland	450,232	771,022	45,023	33,767	28,140
Woomera	988,062	1,750,540	0	478,240	541,404
NorahHead	1,272,045	1,921,794	0		
Townsville	1,034,635	1,511,772	0	492,205	612,745
MountGambier	588,135	1,431,982	0	102,284	451,756
MelbourneAirport	542,291	1,092,220	229,137	76,379	206,223
Nuriootpa	987,581	1,510,418	165,980	713,714	755,209
Launceston	233,620	624,485	0		
WaggaWagga	57,327	1,089,217	0	286,636	745,254
MelbourneAirport	127,367	179,142	14,497	91,124	9,320
Launceston	1,044,934	1,818,016	0		
Darwin	187,598	344,286	0	46,900	107,656
Newcastle		573,720	2,226,034		
Melbourne	1,520,261	2,774,935	0	384,644	1,016,560
Dartmoor	0	995,280	26,020	52,041	422,832
Hobart	194,278	498,940	52,985	17,662	141,293
NorahHead	777,550	1,171,246	0		
Katherine	200,520	394,356	0	106,944	
AliceSprings	1,590,896	2,137,462	0	468,485	0
CoffsHarbour	879,282	1,240,765	87,928		

```
print(xtable(ds,
             digits = 0,
             caption="Selected observations from \\textbf{weatherAUS}."),
      include.rownames=FALSE)
```

```
print(xtable(ds,
             digits = 0,
             caption="Selected observations from \\textbf{weatherAUS}.",
             label="MyTable"),
      include.rownames=FALSE)
```

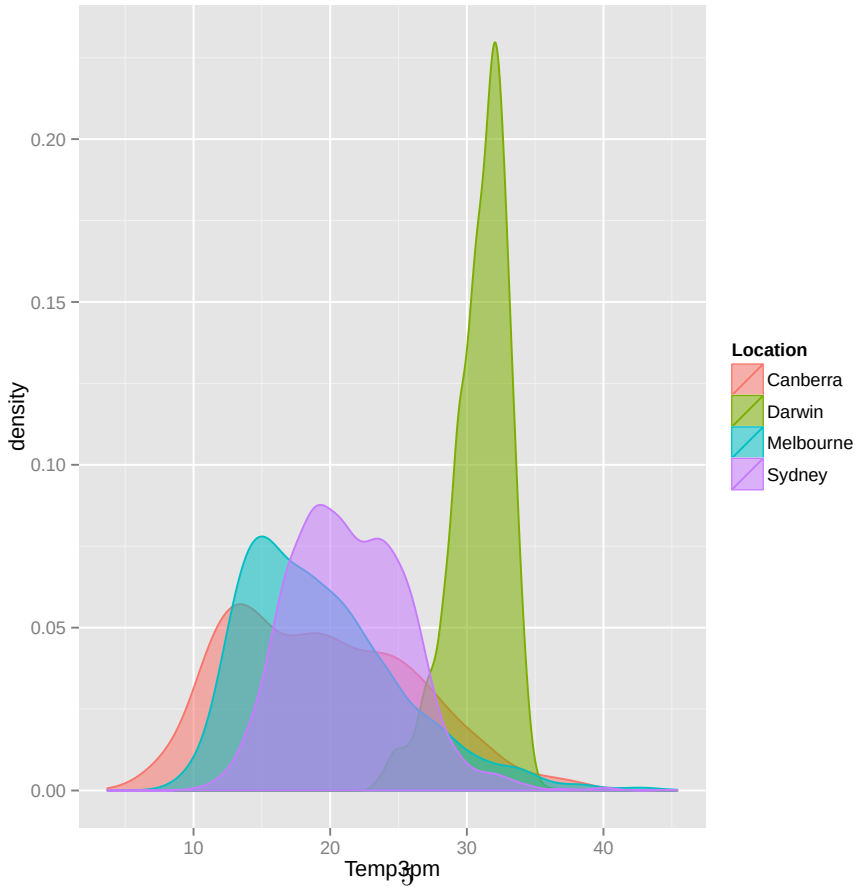
Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine
Portland	8	14	1	1	0
Woomera	22	39	0	11	12
NorahHead	14	21	0		
Townsville	21	30	0	10	12
MountGambier	7	17	0	1	5
MelbourneAirport	7	14	3	1	3
Nuriootpa	12	18	2	9	9
Launceston	5	14	0		
WaggaWagga	1	15	0	4	10
MelbourneAirport	12	17	1	9	1
Launceston	12	21	0		
Darwin	18	32	0	4	10
Newcastle		20	78		
Melbourne	17	30	0	4	11
Dartmoor	0	15	0	1	6
Hobart	4	11	1	0	3
NorahHead	16	24	0		
Katherine	15	30	0	8	
AliceSprings	16	22	0	5	0
CoffsHarbour	18	25	2		

Table 1: Selected observations from **weatherAUS**.

```
print(xtable(ds,
             digits=0,
             caption=paste("Here we include a cample f \\LaTeX{}",
                           "symbols that can be included in the string",
                           "for example",
                           "$\\alpha$ $\\longrightarrow$ $\\wp$.",
                           Sys.time()),
             lable="SymbolCaption"),
      include.rownames=FALSE)
```

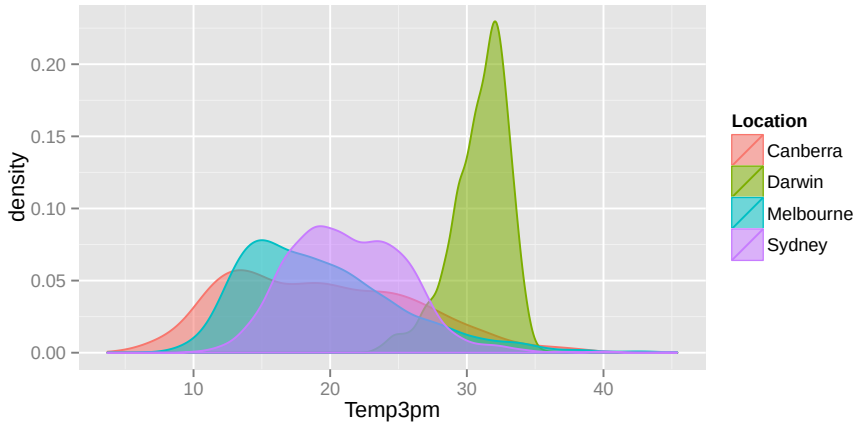
Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine
Portland	8	14	1	1	0
Woomera	22	39	0	11	12
NorahHead	14	21	0		
Townsville	21	30	0	10	12
MountGambier	7	17	0	1	5
MelbourneAirport	7	14	3	1	3
Nuriootpa	12	18	2	9	9
Launceston	5	14	0		
WaggaWagga	1	15	0	4	10
MelbourneAirport	12	17	1	9	1
Launceston	12	21	0		
Darwin	18	32	0	4	10
Newcastle		20	78		
Melbourne	17	30	0	4	11
Dartmoor	0	15	0	1	6
Hobart	4	11	1	0	3
NorahHead	16	24	0		
Katherine	15	30	0	8	
AliceSprings	16	22	0	5	0
CoffsHarbour	18	25	2		

Table 2: Selected observations from **weatherAUS**.



Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine
Portland	8	14	1	1	0
Woomera	22	39	0	11	12
NorahHead	14	21	0		
Townsville	21	30	0	10	12
MountGambier	7	17	0	1	5
MelbourneAirport	7	14	3	1	3
Nuriootpa	12	18	2	9	9
Launceston	5	14	0		
WaggaWagga	1	15	0	4	10
MelbourneAirport	12	17	1	9	1
Launceston	12	21	0		
Darwin	18	32	0	4	10
Newcastle		20	78		
Melbourne	17	30	0	4	11
Dartmoor	0	15	0	1	6
Hobart	4	11	1	0	3
NorahHead	16	24	0		
Katherine	15	30	0	8	
AliceSprings	16	22	0	5	0
CoffsHarbour	18	25	2		

Table 3: Here we include a cample f LaTeX symbols that can be included in the string for example $\alpha \longrightarrow \wp$. 2015-08-25 19:46:16



6 Model Building

Include all models built and parameters tried. Include R code and model evaluations.

7 Deployment

Choose the model to deploy and export it, perhaps as PMML.