

MyFirstRMarkdowndocument

B199291

26 June, 2022

```
#Loading NHSRdatasets and required packages.
```

```
library(NHSRdatasets)
library(tidyverse)
library(here)
library(knitr)
library(scales)
library(lubridate)
library(caret)
library(ggplot2)
```

```
#Loading ae_attendances Dataset
```

```
#Load the ae_attendances data.
```

```
data(ae_attendances)
```

Let's have a look at the ae_attendances data

```
ae<-ae_attendances
class(ae)
```

```
## [1] "tbl_df"      "tbl"        "data.frame"
```

```
glimpse(ae)
```

```
## Rows: 12,765
## Columns: 6
## $ period      <date> 2017-03-01, 2017-03-01, 2017-03-01, 2017-03-01, 2017-03-0~
## $ org_code    <fct> RF4, RF4, RF4, R1H, R1H, R1H, AD913, RYX, RQM, RQM, RJ6, R~
## $ type        <fct> 1, 2, other, 1, 2, other, other, other, 1, other, 1, other~
## $ attendances <dbl> 21289, 813, 2850, 30210, 807, 11352, 4381, 19562, 17414, 7~
## $ breaches   <dbl> 2879, 22, 6, 5902, 11, 136, 2, 258, 2030, 86, 1322, 140, 0~
## $ admissions  <dbl> 5060, 0, 0, 6943, 0, 0, 0, 0, 3597, 0, 2202, 0, 0, 0, 3360~
```

I looked the ae_attendances data review and its class using class function and glimpse function from tidyverse. The data frame has 12,765 rows of data and six columns of different variables with different classes. I can see period (date variable), org_code and type (factor variable), attendances, breaches and admissions as numeric(double precision)variables. As per my intention, I will use type(factor) variable to subset into new data set.

Missing data checking

```
ae %>%
  map(is.na) %>%
```

```
map(sum)
```

```
## $period
## [1] 0
##
## $org_code
## [1] 0
##
## $type
## [1] 0
##
## $attendances
## [1] 0
##
## $breaches
## [1] 0
##
## $admissions
## [1] 0
```

Just to make sure that there is no missing data in the table and the data is complete.

Let's add an index link column to ae_attendances data

For DCT development and training and testing dataset separation, i will add index ref column.

```
ae <- rowid_to_column(ae, "Index")
glimpse(ae)

## Rows: 12,765
## Columns: 7
## $ Index      <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,~
## $ period     <date> 2017-03-01, 2017-03-01, 2017-03-01, 2017-03-01, 2017-03-0~
## $ org_code   <fct> RF4, RF4, RF4, R1H, R1H, R1H, AD913, RYX, RQM, RQM, RJ6, R~
## $ type       <fct> 1, 2, other, 1, 2, other, other, other, 1, other, 1, other~
## $ attendances <dbl> 21289, 813, 2850, 30210, 807, 11352, 4381, 19562, 17414, 7~
## $ breaches  <dbl> 2879, 22, 6, 5902, 11, 136, 2, 258, 2030, 86, 1322, 140, 0~
## $ admissions <dbl> 5060, 0, 0, 6943, 0, 0, 0, 0, 3597, 0, 2202, 0, 0, 0, 3360~
write_csv(ae, here("RefData", "ae_attendances.csv"))
```

Let's subset my raw data first into type 1 org only data

```
ae_type1 <- subset(ae, type=='1')
unique(ae_type1$type)
```

```
## [1] 1
## Levels: 1 2 other
```

```
unique(ae_type1$org_code)
```

```
## [1] RF4 R1H RQM RJ6 RVR RJ1 RQX RYJ RJZ RAX RJ2 R1K RAP RAL RJ7 RAS RKE RRV
## [19] RDD RC1 RQ3 RJF RGT RFS RDE RTG RWH RLT RR1 RQQ RGQ RGP RNQ RC9 RQ8 RD8
## [37] RM1 RGN RNS RX1 RXK RK5 RXW RJC RAJ RNA RQW RCX RL4 RWD RRK RKB RWE RJE
## [55] RBK RWG RGR RWP RLQ REM RCF RBS RFF RXL RMC RAE RWY RW3 RLN RJR RXP RP5
```

```
## [73] RJN RXR RR7 RCD RWA RXN RR8 RBT RXF RNL RVW RJL RTF RW6 RQ6 RM3 RCU RHQ
## [91] RTR RE9 RVY RBN RWJ RMP RTD RFR RM2 RTX RWW RBL RRF RCB RTK RXH RXQ RN7
## [109] RBD RVV RXC RDU RTE RN3 RN5 R1F RWF RPA RVJ RBZ RTH RK9 RD3 RHU RHW REF
## [127] RH8 RA2 RD1 RNZ RTP RBA RDZ RA9 RHM RA7 RYR RA3 RA4 ROA
## 274 Levels: 8J094 AAH AC008 AD913 AF002 AF003 AJN ATQ02 AXG AXT02 ... Y04538
```

Let's tabulate Type1 hospital data and save for my upcoming works

```
ae_type1 %>%
  mutate_at(vars(period), format, "%b-%y") %>%
  mutate_at(vars(attendances, breaches, admissions), comma) %>%
  head(10) %>%
  kable()
```

Index	period	org_code	type	attendances	breaches	admissions
1	Mar-17	RF4	1	21,289.0	2,879.0	5,060.0
4	Mar-17	R1H	1	30,210.0	5,902.0	6,943.0
9	Mar-17	RQM	1	17,414.0	2,030.0	3,597.0
11	Mar-17	RJ6	1	6,654.0	1,322.0	2,202.0
15	Mar-17	RVR	1	12,649.0	473.0	3,360.0
17	Mar-17	RJ1	1	12,385.0	2,092.0	3,181.0
21	Mar-17	RQX	1	10,300.0	635.0	1,684.0
23	Mar-17	RYJ	1	10,020.0	2,632.0	3,270.0
26	Mar-17	RJZ	1	17,810.0	4,153.0	4,477.0
29	Mar-17	RAX	1	9,239.0	968.0	1,839.0

```
write_csv(ae_type1, here("WorkingData", "ae_type1.csv"))
```

Getting Total Attendances of all Type 1 hospitals

```
Type1_attendances <- ae_type1 %>%
  group_by(org_code) %>%
  summarise_at(vars(attendances, breaches), sum)
glimpse(Type1_attendances)
```

```
## Rows: 140
## Columns: 3
## $ org_code    <fct> ROA, R1F, R1H, R1K, RA2, RA3, RA4, RA7, RA9, RAE, RAJ, RAL~
## $ attendances <dbl> 383915, 134308, 998934, 427656, 214089, 150375, 153076, 33~
## $ breaches   <dbl> 84716, 27515, 201171, 130268, 20239, 25711, 5931, 54914, 3~
```

Getting Type1 hospitals with attendances in descending order and saving for record

```
Type1_att_descending <- Type1_attendances[order(-Type1_attendances$attendances),]
glimpse(Type1_att_descending)
```

```
## Rows: 140
## Columns: 3
## $ org_code    <fct> R1H, RW6, RDU, RF4, RXF, RAL, RQM, RR8, RJZ, RWE, RJ2, RRK~
## $ attendances <dbl> 998934, 811852, 719006, 697635, 688461, 688082, 623161, 61~
## $ breaches   <dbl> 201171, 168620, 77526, 150318, 109953, 103247, 52560, 1111~
```

```
write_csv(Type1_att_descending, here("WorkingData", "Type1_attendances_descending.csv"))
```

Subsetting new df with top 5 hospitals

```
Top5_attendance <- head(Type1_att_descending,5)
write_csv(Top5_attendance, here("WorkingData", "Top5_attendance.csv"))
```

Tabulating Top 5 most visited hospitals

```
Top5_attendance %>%
  mutate_at(vars(attendances, breaches), comma) %>%
  kable()
```

org_code	attendances	breaches
R1H	998,934	201,171
RW6	811,852	168,620
RDU	719,006	77,526
RF4	697,635	150,318
RXF	688,461	109,953

Creating new dataset of top 5 hospitals for creating visual

```
Top5_visualready <- filter(ae_type1, org_code %in% Top5_attendance$org_code)
Top5_visualready$performance <- 1-(Top5_visualready$breaches/Top5_visualready$attendances)
glimpse(Top5_visualready)
```

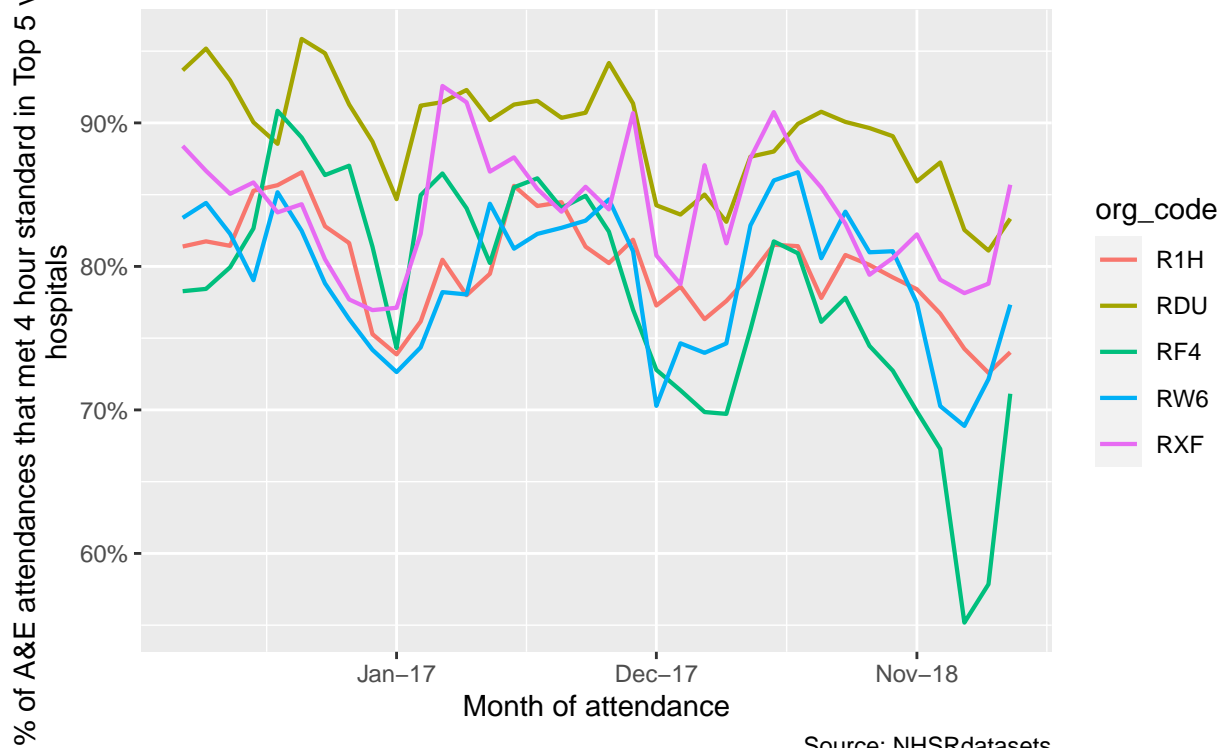
```
## Rows: 180
## Columns: 8
## $ Index      <int> 1, 4, 207, 217, 280, 348, 351, 556, 566, 628, 696, 699, 90~
## $ period     <date> 2017-03-01, 2017-03-01, 2017-03-01, 2017-03-01, 2017-03-0~
## $ org_code   <fct> RF4, R1H, RXF, RW6, RDU, RF4, R1H, RXF, RW6, RDU, RF4, R1H~
## $ type       <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ attendances <dbl> 21289, 30210, 20398, 22930, 20403, 17917, 25955, 17442, 20~
## $ breaches  <dbl> 2879, 5902, 1515, 4997, 1745, 2694, 6186, 3092, 5159, 1527~
## $ admissions <dbl> 5060, 6943, 5047, 5328, 6054, 4208, 5820, 4347, 4778, 5308~
## $ performance <dbl> 0.8647658, 0.8046342, 0.9257280, 0.7820759, 0.9144734, 0.8~
```

```
write_csv(Top5_visualready, here("WorkingData", "Top5_visualready.csv"))
```

Visualising Top 5 hospital breach percentage by time

```
ggplot(Top5_visualready, aes(x=period, y=performance, group=org_code, colour=org_code)) +
  geom_line (size=0.75) + scale_y_continuous(labels = percent) +
  scale_x_date(date_labels = "%b-%y", date_breaks = "11 month")+
  labs(x = "Month of attendance",
       y = "% of A&E attendances that met 4 hour standard in Top 5 visited
       hospitals",
       title = "NHS England accident and emergency (A&E) four hour performance in Top 5
       visited hospitals",
       caption = "Source: NHSRdatasets")+
  theme(plot.title = element_text(hjust = 0.5))
```

NHS England accident and emergency (A&E) four hour performance in Top 5 visited hospitals



Separating provisionalae_type1 data into training and testing sets

Indexing to create train and test data

```
nrow(ae_type1) #to confirm number of rows again

## [1] 4932

prop<-(1-(15/nrow(ae_type1)))
print(prop)

## [1] 0.9969586

set.seed(333)
#Partitioning the raw data into the test and training data.
trainIndex <- createDataPartition(ae_type1$Index, p = prop,
                                   list = FALSE,
                                   times = 1)

head(trainIndex)

##      Resample1
## [1,]         1
## [2,]         2
## [3,]         3
## [4,]         4
## [5,]         5
## [6,]         6
```

```
# All records that are in the trainIndex are assigned to the training data.
ae_type1Train <- ae_type1[ trainIndex,]
nrow(ae_type1Train)
```

```
## [1] 4920
```

There are 12,753 records in your training data. That is a large dataset!

Creating Training Dataset

```
ae_type1Train %>%
  mutate_at(vars(period), format, "%b-%y") %>%
  mutate_at(vars(attendances, breaches), comma) %>%
  head(10) %>%
  kable()
```

Index	period	org_code	type	attendances	breaches	admissions
1	Mar-17	RF4	1	21,289.0	2,879.0	5060
4	Mar-17	R1H	1	30,210.0	5,902.0	6943
9	Mar-17	RQM	1	17,414.0	2,030.0	3597
11	Mar-17	RJ6	1	6,654.0	1,322.0	2202
15	Mar-17	RVR	1	12,649.0	473.0	3360
17	Mar-17	RJ1	1	12,385.0	2,092.0	3181
21	Mar-17	RQX	1	10,300.0	635.0	1684
23	Mar-17	RYJ	1	10,020.0	2,632.0	3270
26	Mar-17	RJZ	1	17,810.0	4,153.0	4477
29	Mar-17	RAX	1	9,239.0	968.0	1839

```
write_csv(ae_type1Train, here("WorkingData", "ae_type1_train.csv"))
```

Creating Testing Dataset

```
ae_type1_test <- ae_type1[-trainIndex,]
nrow(ae_type1_test)
```

```
## [1] 12
```

```
ae_type1TestMarker <- ae_type1_test[1,]
ae_type1TestMarker %>%
  mutate_at(vars(period), format, "%b-%y") %>%
  mutate_at(vars(attendances, breaches), comma) %>%
  head(10) %>%
  kable()
```

Tabulating and saving TestMarker

Index	period	org_code	type	attendances	breaches	admissions
318	Mar-17	RD1	1	6,218	1,379	2103

```
write_csv(ae_type1TestMarker, here("WorkingData", "ae_type1_testmarker.csv"))
```

Saving remaining test records

```
ae_type1_test <- ae_type1_test[2:nrow(ae_type1_test),]
ae_type1_test %>%
  mutate_at(vars(period), format, "%b-%y") %>%
  mutate_at(vars(attendances, breaches), comma) %>%
  head(10) %>%
  kable()
```

Index	period	org_code	type	attendances	breaches	admissions
2826	Jul-16	RGT	1	10,081	2,288	3123
2865	Jul-16	RM1	1	10,303	1,080	2891
3941	Apr-16	RRK	1	9,044	1,195	2744
5431	Dec-17	RJN	1	4,238	1,208	971
5940	Nov-17	RNZ	1	3,940	209	1131
6727	Aug-17	RQ3	1	3,801	86	754
7952	May-17	RBL	1	8,270	2,120	2491
8427	Mar-19	R1K	1	12,753	3,178	6522
10489	Oct-18	RPA	1	7,710	1,750	2065
12390	May-18	RA3	1	4,390	348	1090

```
write_csv(ae_type1_test, here("WorkingData", "ae_type1_test.csv"))
```