Loading NHSRDatasets for Exploratory Analysis and Data Capture Tool

B199291

27 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

```
data(ae_attendances)
```

Let's have a look at the ae_attendances data

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

I looked the ae_attedances 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 numerice(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)
```

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

Let's tablulate 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()
write_csv(ae_type1, here("WorkingData", "ae_type1.csv"))
```

Getting Total Attendances of all Type 1 hopsitals

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

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)
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()
```

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)
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))
```

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
prop<-(1-(15/nrow(ae_type1)))
print(prop)
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)
# All records that are in the trainIndex are assigned to the training data.
ae_type1Train <- ae_type1[ trainIndex,]
nrow(ae_type1Train)
```

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

Creating Testing Dataset

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

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

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

Tabulating and saving TestMarker

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