datamining-project-impuation

DEEP SAGAR KARKI

4/1/2022

```
library(magrittr)
```

```
## Warning: package 'magrittr' was built under R version 4.1.3
```

```
library(tidyverse)
```

```
## -- Attaching packages -------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.3      v dplyr   1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.1.1
```

```
## Warning: package 'readr' was built under R version 4.1.3
```

```
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x tidyr::extract()   masks magrittr::extract()
## x dplyr::filter()    masks stats::filter()
## x dplyr::lag()       masks stats::lag()
## x purrr::set_names() masks magrittr::set_names()
```

```
data <- read.csv("C:/Users/14193/Downloads/Life Expectancy Data.csv") %>%
        filter(Year==2014)
```

checking for missing values

```
sapply(data, anyNA)
```

```
##               Country                    Year
##                 FALSE                   FALSE
##                Status         Life.expectancy
##                 FALSE                   FALSE
##        Adult.Mortality            infant.deaths
##                 FALSE                   FALSE
```

```
##                         Alcohol          percentage.expenditure
##                            TRUE                           FALSE
##                     Hepatitis.B                         Measles
##                            TRUE                           FALSE
##                             BMI                under.five.deaths
##                            TRUE                           FALSE
##                           Polio                Total.expenditure
##                           FALSE                            TRUE
##                       Diphtheria                         HIV.AIDS
##                           FALSE                           FALSE
##                             GDP                       Population
##                            TRUE                            TRUE
##              thinness..1.19.years              thinness.5.9.years
##                            TRUE                            TRUE
## Income.composition.of.resources                        Schooling
##                            TRUE                            TRUE
```

imputing missing data by k-nn

```r
remotes::install_github("cran/DMwR")
```

```
## Skipping install of 'DMwR' from a github remote, the SHA1 (6fd4f0cd) has not changed since last insta
##   Use 'force = TRUE' to force installation
```

```r
library(DMwR)
```

```
## Loading required package: lattice
```

```
## Loading required package: grid
```

```
## Registered S3 method overwritten by 'quantmod':
##   method             from
##   as.zoo.data.frame zoo
```

```r
knnOutput <- knnImputation(data[,-c(1,2,3)],k=10,scale=T,)# perform knn imputation except on country, s

# knn output is our data with imputed values


## combining the remaining 3 columns


full_data=cbind(data[,c(1,2,3)],knnOutput)


# checking for any missing values
anyNA(full_data)
```

```
## [1] FALSE
```

```
# no missing values found.
```