MBTI dataset transformation and analysis

Vedran Moškov, Lucija Runjić, Borna Josipović, Lana Bartolović

2024-01-13

```
rm(list = ls())
```

Učitavanje i uređivanje podatkovnog skupa

Učitavanje i proučavanje podatkovnog skupa

Učitavamo podatkovni skup u varijablu "dataset".

```
dataset <- read_csv("../data/MBTI.csv")
```

```
## New names:
## Rows: 97 Columns: 21
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (4): SEX, ACTIVITY LEVEL, MBTI, POSTURE dbl (17): ...1, S No, AGE, HEIGHT,
## WEIGHT, PAIN 1, PAIN 2, PAIN 3, PAIN 4, E...
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

Proučavamo podatkovni skup kako bi ga znali urediti na način da nam je lakše raditi s njim kasnije.

```
head(dataset)
```

```
## # A tibble: 6 x 21
##    ...1 `S No`   AGE HEIGHT WEIGHT SEX    `ACTIVITY LEVEL` `PAIN 1` `PAIN 2`
##   <dbl>  <dbl> <dbl>  <dbl>  <dbl> <chr>  <chr>               <dbl>    <dbl>
## 1     0      1    53     62    125 Female Low                   0        0
## 2     1      2    52     69    157 Male   High                  7        8
## 3     2      3    30     69    200 Male   High                  0        0
## 4     3      4    51     66    175 Male   Moderate              9.5      9.5
## 5     4      5    45     63    199 Female Moderate              4        5
## 6     5      6    68     74    182 Male   Low                   0        2.5
## # i 12 more variables: `PAIN 3` <dbl>, `PAIN 4` <dbl>, MBTI <chr>, E <dbl>,
## #   I <dbl>, S <dbl>, N <dbl>, T <dbl>, F <dbl>, J <dbl>, P <dbl>,
## #   POSTURE <chr>
```

```
tail(dataset)
```

```
## # A tibble: 6 x 21
##    ...1 `S No`   AGE HEIGHT WEIGHT SEX    `ACTIVITY LEVEL` `PAIN 1` `PAIN 2`
##   <dbl>  <dbl> <dbl>  <dbl>  <dbl> <chr>  <chr>               <dbl>    <dbl>
## 1    91     92    16     69    130 Female Moderate                5        0
## 2    92     93    16     58    100 Male   Moderate                0        0
## 3    93     94    45     62    134 Female Moderate                0        4
## 4    94     95    43     69    188 Male   Moderate                2        0
## 5    95     96    28     67    180 Female Low                     0        0
## 6    96     97    43     69    188 Male   Moderate                4        0
## # i 12 more variables: `PAIN 3` <dbl>, `PAIN 4` <dbl>, MBTI <chr>, E <dbl>,
## #   I <dbl>, S <dbl>, N <dbl>, T <dbl>, F <dbl>, J <dbl>, P <dbl>,
## #   POSTURE <chr>
```

```
glimpse(dataset)
```

```
## Rows: 97
## Columns: 21
## $ ...1             <dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,~
## $ `S No`           <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16~
## $ AGE              <dbl> 53, 52, 30, 51, 45, 68, 62, 65, 66, 58, 61, 33, 48, 5~
## $ HEIGHT           <dbl> 62, 69, 69, 66, 63, 74, 68, 61, 67, 69, 67, 62, 64, 6~
## $ WEIGHT           <dbl> 125, 157, 200, 175, 199, 182, 263, 143, 180, 165, 210~
## $ SEX              <chr> "Female", "Male", "Male", "Male", "Female", "Male", "~
## $ `ACTIVITY LEVEL` <chr> "Low", "High", "High", "Moderate", "Moderate", "Low",~
## $ `PAIN 1`         <dbl> 0.0, 7.0, 0.0, 9.5, 4.0, 0.0, 7.0, 0.0, 0.5, 0.0, 5.0~
## $ `PAIN 2`         <dbl> 0.0, 8.0, 0.0, 9.5, 5.0, 2.5, 10.0, 9.0, 3.5, 7.5, 0.~
## $ `PAIN 3`         <dbl> 0.0, 5.0, 0.0, 9.5, 2.0, 1.5, 10.0, 5.0, 0.5, 7.0, 0.~
## $ `PAIN 4`         <dbl> 0.0, 3.0, 0.0, 1.5, 2.0, 0.0, 10.0, 10.0, 9.5, 3.0, 9~
## $ MBTI             <chr> "ESFJ", "ISTJ", "ESTJ", "ISTJ", "ENFJ", "ISFP", "ISTP~
## $ E                <dbl> 0.9084579, -0.6045853, 0.4727891, -0.6045853, 0.34875~
## $ I                <dbl> -1.0968036, 0.4727891, -0.6045853, 0.4727891, -0.4727~
## $ S                <dbl> -0.06968492, -0.28221615, -0.13971030, 0.21042839, 0.~
## $ N                <dbl> -0.6744898, -0.4307273, -0.5894558, -1.0853249, -0.96~
## $ T                <dbl> -0.3186394, 1.1503494, 0.3186394, 0.1046335, -0.31863~
## $ F                <dbl> 0.1046335, -1.1503494, -0.3186394, -0.1046335, 0.3186~
## $ J                <dbl> 0.78103381, 0.16421078, 0.05451891, 0.93881432, 0.511~
## $ P                <dbl> -0.93881432, -0.27592106, -0.16421078, -1.12433823, -~
## $ POSTURE          <chr> "A", "B", "A", "D", "A", "D", "B", "D", "C", "D", "B"~
```

Uređivanje podataka podatkovnog skupa

Faktoriziramo određene stupce "SEX", "ACTIVITY LEVEL", "MBTI" kako bismo kasnije mogli lakše grupirati podatke i bolje ih analizirati

```
dataset$SEX <- as.factor(dataset$SEX)
dataset$`ACTIVITY LEVEL` <- as.factor(dataset$`ACTIVITY LEVEL`)
dataset$MBTI <- as.factor(dataset$MBTI)
```

Analiza podatkovnog skupa