

descriptive table

bf2506

2023-05-07

```
library(tidyverse)
library(caret)
library(summarytools)
```

1. Load Data

```
load("final_data.RData")
final_data = final_data %>%
  mutate(binary_recovery_time = as.factor(binary_recovery_time))
levels(final_data$binary_recovery_time) = c("low", "high")
```

2. Train/test split

```
set.seed(2)
training_rows <- createDataPartition(final_data$binary_recovery_time,
                                     p = 0.8,
                                     list = F)
```

```
training_set <- final_data[training_rows,]
```

```
st_options(plain.ascii = FALSE,
            style = "rmarkdown",
            dfSummary.silent = TRUE,
            footnote = NA,
            subtitle.emphasis = FALSE)
```

```
dfSummary(training_set[, -1])
```

Data Frame Summary

training_set
Dimensions: 2888 x 16
Duplicates: 0

| No | Variable | Stats / Values | Freqs (% of Valid) | Graph | Valid | Missing |
|----|--------------------------|--|---|---|------------------|-------------|
| 1 | age [numeric] | Mean (sd) : 60.1 (4.4) min < med < max: 42 < 60 < 79 IQR (CV) : 6 (0.1) | 33 distinct values | : . : : : : . : : : : : : | 2888 (100.0%) | 0 (0.0%) |
| 2 | gender [factor] | 1. 0 2. 1 | 1475 (51.1%) 1413 (48.9%) | IIIIIIII IIIIIIII | 2888 (100.0%) | 0 (0.0%) |
| 3 | race [factor] | 1. 1 2. 2 3. 3 4. 4 | 1840 (63.7%) 151 (5.2%) 599 (20.7%) 298 (10.3%) | IIIIIIIIII I III II | 2888 (100.0%) | 0 (0.0%) |
| 4 | smoking [factor] | 1. 0 2. 1 3. 2 | 1784 (61.8%) 832 (28.8%) 272 (9.4%) | IIIIIIIIII IIII I | 2888 (100.0%) | 0 (0.0%) |
| 5 | height [numeric] | Mean (sd) : 169.9 (6.1) min < med < max: 151.2 < 170.1 < 195.9 IQR (CV) : 7.8 (0) | 320 distinct values | . : : : : : : : : : : : : : : : . | 2888 (100.0%) | 0 (0.0%) |
| 6 | weight [numeric] | Mean (sd) : 80.1 (7) min < med < max: 57.9 < 80.1 < 105 IQR (CV) : 9.5 (0.1) | 349 distinct values | . : : : : : : . : : : : : . : : : : : . | 2888 (100.0%) | 0 (0.0%) |
| 7 | bmi [numeric] | Mean (sd) : 27.8 (2.7) min < med < max: 19.9 < 27.7 < 38.9 IQR (CV) : 3.5 (0.1) | 161 distinct values | . : : : . : : . : : : : . : : : : : . | 2888 (100.0%) | 0 (0.0%) |
| 8 | hypertension [factor] | 1. 0 2. 1 | 1525 (52.8%) 1363 (47.2%) | IIIIIIII IIIIIIII | 2888 (100.0%) | 0 (0.0%) |
| 9 | diabetes [factor] | 1. 0 2. 1 | 2445 (84.7%) 443 (15.3%) | IIIIIIIIIIIIII III | 2888 (100.0%) | 0 (0.0%) |
| 10 | sbp [numeric] | Mean (sd) : 130 (7.8) min < med < max: 105 < 130 < 158 IQR (CV) : 10 (0.1) | 54 distinct values | . : : : : : : . : : : . : : : : : . | 2888 (100.0%) | 0 (0.0%) |
| 11 | ldl [numeric] | Mean (sd) : 110.2 (19.6) min < med < max: 32 < 110 < 175 IQR (CV) : 26 (0.2) | 118 distinct values | . : : : : : : : : : . : : : : . | 2888 (100.0%) | 0 (0.0%) |
| 12 | vaccine [factor] | 1. 0 2. 1 | 1177 (40.8%) 1711 (59.2%) | IIIIII IIIIIIIIII | 2888 (100.0%) | 0 (0.0%) |
| 13 | severity [factor] | 1. 0 2. 1 | 2617 (90.6%) 271 (9.4%) | IIIIIIIIIIIIIIII I | 2888 (100.0%) | 0 (0.0%) |
| 14 | study [factor] | 1. A 2. B 3. C | 582 (20.2%) 1734 (60.0%) 572 (19.8%) | III IIIIIIIIII III | 2888 (100.0%) | 0 (0.0%) |

| No | Variable | Stats / Values | Freqs (% of Valid) | Graph | Valid | Missing |
|----|----------------------------------|---|-----------------------------|-----------------------------------|------------------|-------------|
| 15 | recovery_time [numeric] | Mean (sd) : 43 (29.5) min < med < max: 2 < 39 < 365 IQR (CV) : 22 (0.7) | 152 distinct values | : . : : : : : : : : . | 2888 (100.0%) | 0 (0.0%) |
| 16 | binary_recovery_time [factor] | 1. low 2. high | 896 (31.0%) 1992 (69.0%) | IIIIII IIIIIIIIIIII | 2888 (100.0%) | 0 (0.0%) |