# Module 2 - Factor

Rsquared Academy

2021-03-05

## Import Data & Load Libraries

```
analytics <- readRDS("analytics_raw.rds")
library(forcats)
```

1. Check whether the below variables are factor

    - device
    - page_depth
    - landing_page

```
is.factor(analytics$device)
```

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

```
is.factor(analytics$page_depth)
```

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

```
is.factor(analytics$landing_page)
```

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

2. Coerce the following variables to type factor

    - device
    - os
    - browser
    - user_type
    - channel
    - gender
    - landing_page
    - exit_page
    - city
    - country
    - user_type

```
analytics$device       <- as.factor(analytics$device)
analytics$os           <- as.factor(analytics$os)
analytics$browser      <- as.factor(analytics$browser)
analytics$user_type    <- as.factor(analytics$user_type)
analytics$channel      <- as.factor(analytics$channel)
analytics$gender       <- as.factor(analytics$gender)
analytics$landing_page <- as.factor(analytics$landing_page)
analytics$exit_page    <- as.factor(analytics$exit_page)
analytics$country      <- as.factor(analytics$country)
analytics$user_rating  <- as.factor(analytics$user_rating)
```

3. Use only the following levels in the `gender` column:

    - `male`
    - `female`

```
table(factor(analytics$gender, levels = c("male", "female")))
```

```
##
##   male female
##  61617  40565
```

4. Include `NA` as a level in the gender column.

```
table(factor(analytics$gender, levels = c("male", "female", NA)))
```

```
##
##   male female
##  61617  40565
```

5. Change label of `NA` to `missing` in the `gender` column.

```
table(factor(analytics$gender,
       levels = c("male", "female", "NA"),
       labels = c("male", "female", "missing")))
```

```
##
##    male  female missing
##   61617   40565       0
```

6. Change the labels of the levels in `user_type` column to

    - `New`
    - `Returning`

```
table(factor(analytics$user_type,
       levels = c("New Visitor", "Returning Visitor"),
       labels = c("new", "returning")))
```

```
##
##       new returning
##    179045     65353
```

7. Check if the `user_rating` column is ordered. If not, coerce it to type ordered factor.

```
# membership testing
is.ordered(analytics$user_rating)
```

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

```
# using ordered
table(ordered(analytics$user_rating, levels = c("1", "2", "3", "4", "5")))
```

```
##
##      1      2      3      4      5
##  14664   4888  12220 158859  53767
```

```
# using factor
table(factor(analytics$user_rating,
       levels  = c("1", "2", "3", "4", "5"),
       ordered = TRUE))
```

```
##
##      1      2      3      4      5
##  14664   4888  12220 158859  53767
```