

Group 13 GLM - Ikea Dataset

Group 13

14/03/2022

The dataset contains information on IKEA furniture from Saudi Arabia, a column has been added indicating if the piece of furniture costs more than 1000 Riyals.

```
##   X   item_id                category price sellable_online other_colors
## 1 1 60406785                Chairs   149             TRUE          No
## 2 2 99221446          Tables & desks 5465             TRUE          No
## 3 3 29221765 Bookcases & shelving units 660             TRUE          No
## 4 4 20403374          Sofas & armchairs 395             TRUE          Yes
## 5 5 89256262                Chairs 1895             TRUE          Yes
## 6 6 50165079          Nursery furniture 375             TRUE          No
##   depth height width thousandRiyals
## 1    44    103    52              No
## 2    NA     NA    NA              Yes
## 3    35    176    86              No
## 4    NA     NA    NA              No
## 5   164    104   117              Yes
## 6    53     93    52              No
```

The summary of the data indicates that there is a lot of NA values in the height, depth and width columns.

```
##           X           item_id           category           price
## Min.      : 1.0   Min.      : 91415   Length:500   Min.      : 3.0
## 1st Qu.:125.8   1st Qu.:20344743   Class :character 1st Qu.: 168.8
## Median :250.5   Median :49284139   Mode  :character Median : 457.0
## Mean    :250.5   Mean    :48073026           Mean    : 991.1
## 3rd Qu.:375.2   3rd Qu.:70414743           3rd Qu.:1245.0
## Max.    :500.0   Max.    :99305158           Max.    :8551.0
##
## sellable_online other_colors           depth           height
## Mode :logical   Length:500   Min.      : 1.00   Min.      : 3.0
## FALSE:1         Class :character 1st Qu.: 37.00   1st Qu.: 68.0
## TRUE :499       Mode  :character Median : 46.00   Median : 83.0
##                                     Mean    : 53.34   Mean    :102.3
##                                     3rd Qu.: 60.00   3rd Qu.:123.8
##                                     Max.    :252.00   Max.    :251.0
##                                     NA's    :191     NA's    :146
##
## width           thousandRiyals
## Min.      : 2.0   Length:500
## 1st Qu.: 56.0   Class :character
## Median : 80.0   Mode  :character
## Mean    :101.1
```

```
## 3rd Qu.:134.2
## Max.    :367.0
## NA's    :80
```

To solve this, I computed the mean of the depth, height and width by the category and replaced the NA's by this value shown below.

```
## # A tibble: 500 x 10
## # Groups:   category [17]
##       X item_id category price sellable_online other_colors depth height width
##   <int>   <int> <chr>    <dbl> <lgl>          <chr>      <dbl> <dbl> <dbl>
## 1     1    60406785 Chairs      149 TRUE          No         44    103    52
## 2     2    99221446 Tables ~    5465 TRUE          No        50.5   65.7   74.1
## 3     3    29221765 Bookcas~    660 TRUE          No         35    176    86
## 4     4    20403374 Sofas &~    395 TRUE          Yes       107.    76.6  162.
## 5     5    89256262 Chairs    1895 TRUE          Yes       164    104   117
## 6     6    50165079 Nursery~    375 TRUE          No         53     93    52
## 7     7    40284191 Sofas &~    150 TRUE          No       107.    15   162.
## 8     8    19291315 Cabinet~    375 TRUE          No         47     92    45
## 9     9     439468 Wardrob~    995 TRUE          No         47    186   120
## 10    10    49252553 Chairs     400 TRUE          No         61     87    56
## # ... with 490 more rows, and 1 more variable: thousandRiyals <chr>
```