Group 13 GLM - Ikea Dataset

Group 13

14/03/2022

Introduction and Cleaning the Data

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>
```

Exploratory Analysis

This section is to explore the data set graphically and gain insight before fitting the Generalised Linear Model (GLM).

This table shows how many pieces of furniture are in each category for the data and is then shown as a barchart. It can be seen that the lowest number of furniture items is for Room dividers and the highest is for Chairs.There is a lot of variability in the number of items in each category.

Table 1:   Total number of items of furniture by category

category	count
Bar furniture	8
Beds	30
Bookcases & shelving units	71
Cabinets & cupboards	47
Café furniture	3
Chairs	74
Chests of drawers & drawer units	13
Children's furniture	14
Nursery furniture	18
Outdoor furniture	33
Room dividers	1
Sideboards, buffets & console tables	4
Sofas & armchairs	51
Tables & desks	67
Trolleys	3
TV & media furniture	27
Wardrobes	36
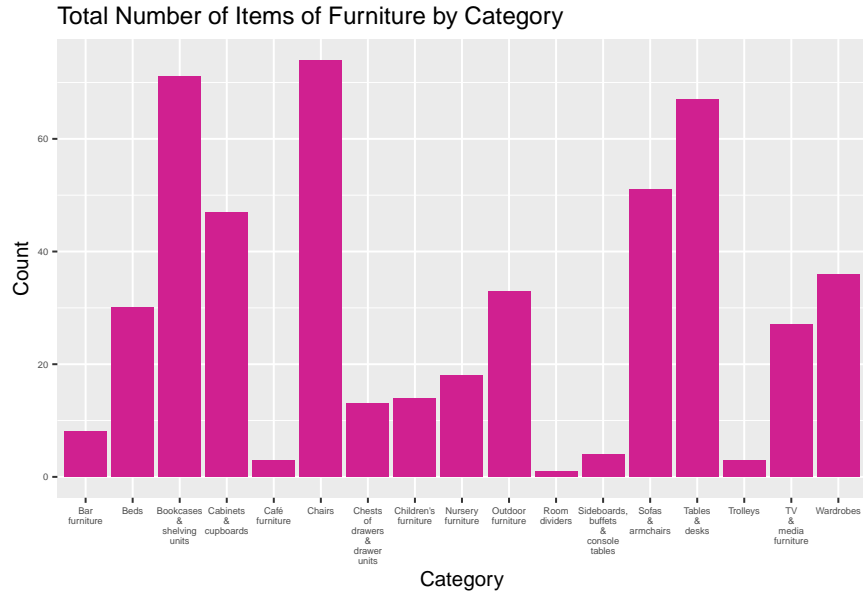
Figure 1:   Total Number of Items of Furniture by Category.

The barchart below is showing the total number in each category by if they are priced above or below 1000 Riyals.It shows that some categories have no items priced greater than 1000 Riyals.
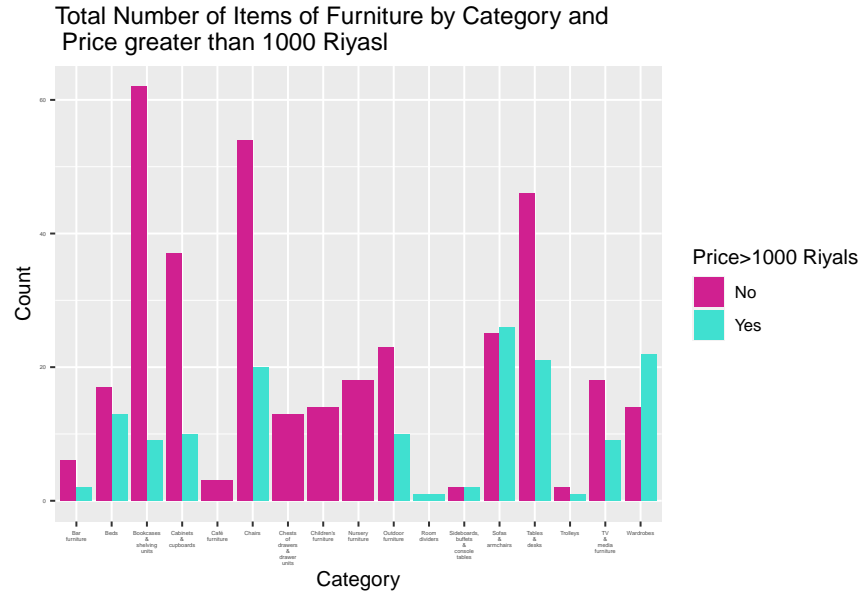
Figure 2: Total Number of Items of Furniture by Category and if they cost greater than 1000 Riyals.

The below box plots explore the relationship between some categorical variables and the price of the item. The first shows some potential outliers in some categories which are much more expensive than other items in that category.
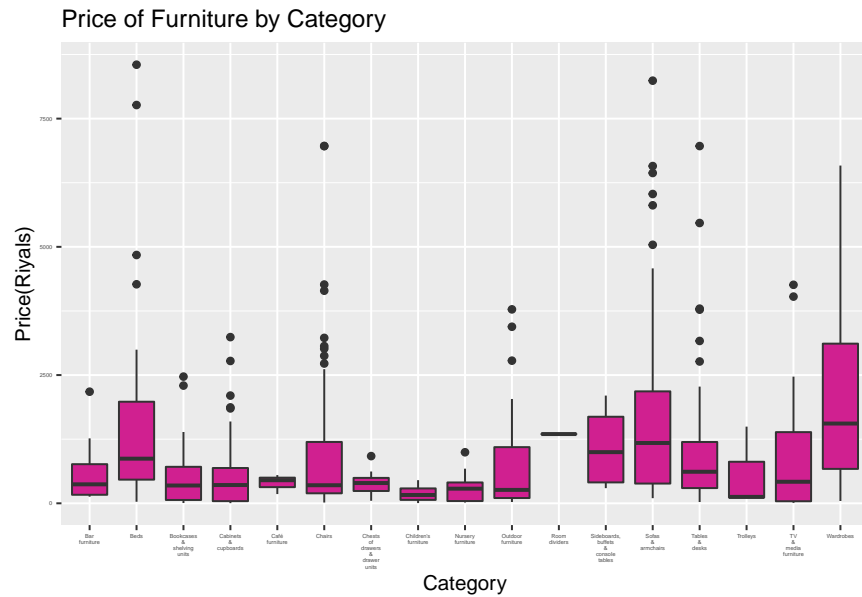

Figure 3: Price of the Item of Furniture by Category.

The box plot below shows whether it is available to buy in other colours and the items price.It appears that the price range for items available in other colours is slightly higher.
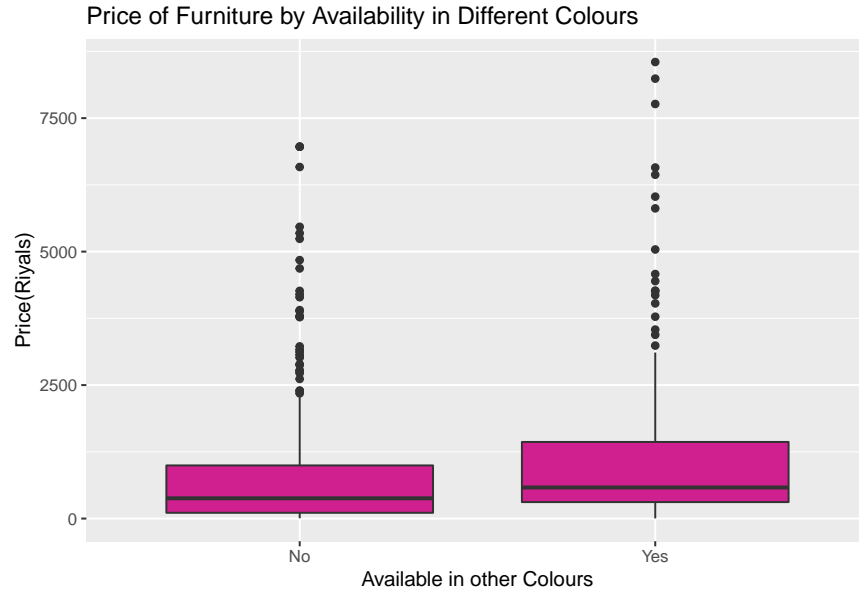
Figure 4:   Price of the Item of Furniture by Availability in other Colours.

The next three plots show the height, width and depth of the item of furniture and whether it costs more than 1000 Riyals. These show that the items that cost more than 1000 Riyals have bigger height, width, depth and general size.

Figure 5:   Height of Item by whether the Price is greater than 1000 Riyals.

Width of Furniture by Price greater than 1000 Riyals

Figure 6: Width of Item by whether the Price is greater than 1000 Riyals.

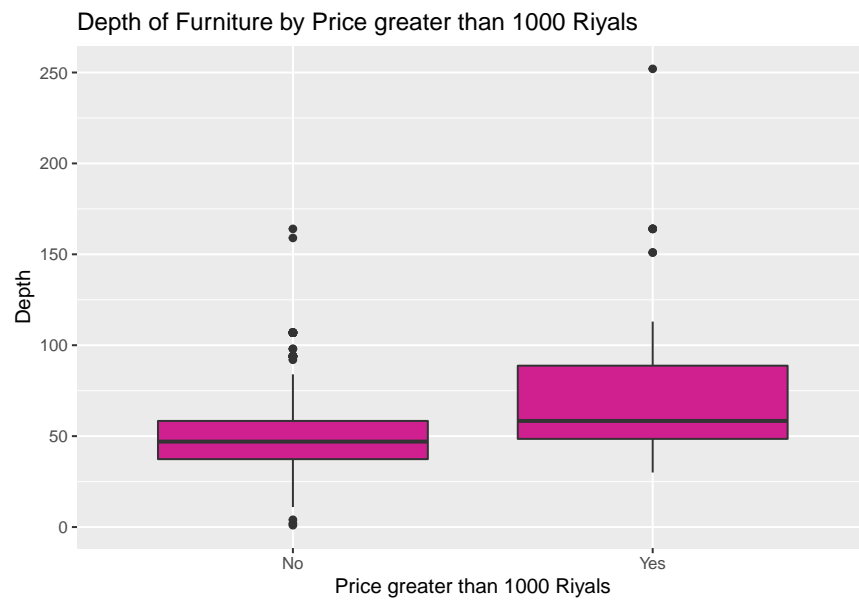Depth of Furniture by Price greater than 1000 Riyals

Figure 7: Depth of Item by whether the Price is greater than 1000 Riyals.