Ikea furnitures EDA

Clinton Oyogo

2020-11-04

**I am loving the tidytuesday challenges cause of the way they stretch my thinking particularly on data visualization skills. Took some time to flex my EDA skills with the TidyTuesday challenge data from Ikea home furnishing company. There are quite a number of things to inspect but I’ll just have a look at a few.**

### Data import

ikea <- readr::read\_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-11-03/ikea.csv')

### Loading libraries

library(dplyr)  
library(plotly)  
library(ggridges)

## How’s the rank of furniture categories with regard to number of designers?

I thought it interesting to see which category of furnitures had the highest number of designers. Its clear from the bar graph below that Tables and desks had the highest count followed closely by Book cases and shelving units.

I love the interactivity of plotly. With a hover on the bars you can see the furniture category and the corresponding number of designers.

ikea\_cat\_des <- ikea %>%  
 dplyr::group\_by(category) %>%  
 dplyr::count(designer) %>%  
 dplyr::summarise(n=sum(n))  
   
   
plot\_ly(ikea\_cat\_des,   
 y = ~reorder(category,n),   
 x = ~n,   
 colors = 'PuRd',  
 showscale = FALSE,  
 showlegend = FALSE,  
 text = ~paste0("Category: ", category, "\n", "Number of designers: ", n)) %>%  
 add\_bars( hoverinfo='text') %>%  
 hide\_colorbar() %>%  
   
 layout(bargap = 0.5,   
 group = ~category ,   
 title = paste("Number of designers per category of furniture"),   
 xaxis = list(title = "Number of designers",tickangle = 45  
   
 ),  
 yaxis=list(title="Category of furniture"),   
 dtick = 1,  
 tick0 = 0,  
 tickmode = "array")

## How’s the comparison of item categories that are sellable online and those that are not?

I love the way a sankey diagram highlights the fact that a large percentage of items are sellable online. As we can note from the diagram, the following items are not sellable online entirely: - cafe furniture,  
- sideboards, buffets and console tables and  
- Room dividers  
whereas for the other categories, only a small percentage of items are not sellable online.

category\_selleable\_online <- ikea %>%  
 dplyr::group\_by(category,sellable\_online) %>%  
 dplyr::count(sellable\_online)  
  
  
 colnames(category\_selleable\_online) <- c("source", "target", "value")  
  
 category\_selleable\_online$source <- stringr::str\_trim(category\_selleable\_online$source)  
 category\_selleable\_online$target <- stringr::str\_trim(category\_selleable\_online$target)  
  
 category\_selleable\_online$target <- paste(category\_selleable\_online$target, " ", sep = "")  
  
 nodes <- data.frame(name=c(as.character(category\_selleable\_online$source), as.character(category\_selleable\_online$target)) %>% unique())  
   
 category\_selleable\_online$IDsource = match(category\_selleable\_online$source, nodes$name)-1  
 category\_selleable\_online$IDtarget = match(category\_selleable\_online$target, nodes$name)-1  
   
   
 sankey\_plot <- plot\_ly(category\_selleable\_online,  
 type = "sankey",  
 node = list(  
 label = nodes$name,  
 pad = 15,  
 thickness = 15,  
 line = list(  
 color = "black",  
 width = 0.5  
 )  
 ),  
   
 link = list(  
 source = category\_selleable\_online$IDsource,  
 target = category\_selleable\_online$IDtarget,  
 value = category\_selleable\_online$value,  
 label = nodes$name  
 )  
 ) %>%  
 layout(  
 title = "Category of items and their online sale status",  
 xaxis = list(showgrid = F, zeroline = F),  
 yaxis = list(showgrid = F, zeroline = F),  
 margin = list( t = 50)  
 )  
 sankey\_plot

## How do the furniture categories compare in terms of Mean price?

ikea\_cat\_price <- ikea %>%  
 dplyr::group\_by(category) %>%  
 dplyr::summarise(mean\_price=mean(price))  
   
   
plot\_ly(ikea\_cat\_price,   
 y = ~reorder(category,mean\_price),   
 x = ~mean\_price,   
 colors = 'PuRd',  
 showscale = FALSE,  
 showlegend = FALSE,  
 text = ~paste0("Category: ", category, "\n", "Mean price: ", mean\_price)) %>%  
 add\_bars( hoverinfo='text') %>%  
 hide\_colorbar() %>%  
   
 layout(bargap = 0.5,   
 group = ~category ,   
 title = paste("Mean price per category of furniture"),   
 xaxis = list(title = "Mean price (Saudi Riyals)",tickangle = 45),  
 yaxis=list(title="Category of furniture"),   
 dtick = 1,  
 tick0 = 0,  
 tickmode = "array")

## Checking the price distributions of the furniture categories

item\_price\_distribution <- ikea %>%  
 dplyr::group\_by(category, price)  
  
  
ggplot(item\_price\_distribution,aes(x = price, y = reorder(category, price))) +  
 geom\_boxplot() +  
 labs(title = "Price distribution per category", y = "Item categories", x = "Price in Saudi Riyals") +  
 theme(plot.title = element\_text(hjust = 0.5, face = "bold"))

