# Report

## A short overview:

**Ikea.com** is a website for the popular multinational group of Swedish companies (IKEA). This website provides many services to the user letting him to take a look of the products and different categories , also it provides shopping online service which is useful these days.

## Data we collect:

### Rows:

First 100 rows are observations for items in Bed category, the second 100 rows are observations for items in Wardrobe (storage and organization) category, last 100 rows are observations for items in Chairs category.

### Columns:

1. Category: Category of the item
2. Name: Name of the item
3. Price: Price of the item
4. Description: Basic Description about the item
5. Measurement: Measurements and Size of its side and other lengths information
6. SKU: A primary Key also known as stock-keeping unit that is unique for each item

## Dataset description:

For the 'Price' feature, the data considers numeric (float) datatype, it has mean of 746.95 and standard deviation of 676.5 whereas the minimum price is 16 SAR and the maximum is 2800 SAR, the median is 559 SAR.

Moreover, 'Category' feature is a categorical datatype (nominal) since it does not have ordering. It has 147 unique values and 13 frequencies. Also 'Name' feature has 83 unique value and 28 frequencies.

In Addition, 'Description' feature has 101 unique values and 19 frequencies, lastly 'Measurement' feature has 150 unique values with 6 frequencies only.

## Challenges we faced in data collection and how we overcome it:

There was a problem regarding bot detector of the website, it prevents us from collecting more data and redirects us to the reCAPTCHA, to solve this problem we used the wait methodology by using sleep( ) function. Also we face challenge in identifying html classes for the categories, but we search and inspect the elements carefully until we solve the problem.

# Questions:

What is the most interesting thing you can discover from the data?

This data has a primary key i.e. SKU that can be used to create a quality database and link everything in a proper manner.

How many item are there for each category?

100 items for each category and different sub categories

Which Category has cheapest item and the most expensive one?

Most Cheapest Category is "Products/Furniture/Chairs/Chair pads/STRÃ…FLY Chair pad"

Most Expensive Category is "/Furniture/Beds/Double beds/TUFJORD Upholstered bed frame"

## Is there a relationship between 'Price' and 'Category'?

Yes , different categories have different scale of prices.

We conduct two plots:

1- **Scatter plot** representing correlation between them and as it shown different categories have different price scales.
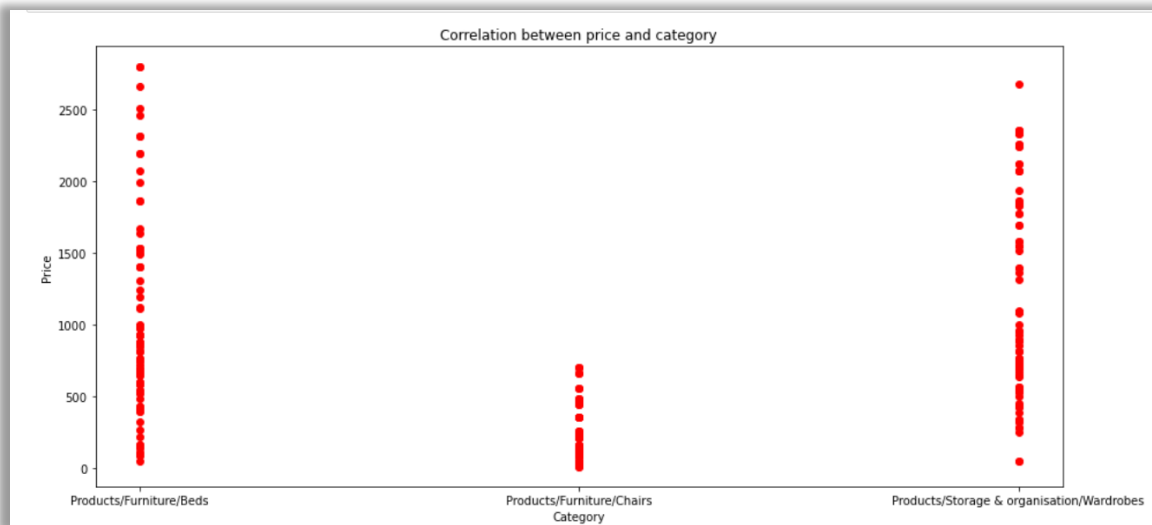


2- **Bar plot** which also represents relationship between the categorical feature 'Category' and numerical feature 'Price', and as shown below we can insure that different categories have different price scales. as a finding from the visualization we can see that beds has higher prices than other categories.