# Introducción

En este documento se denotan los cambios necesarios que se le hacen a la base de datos (BD). Estos cambios se resumen en dos: 1- correcciones de errores como campos nulos o campos en clasificaciones equivocadas, y 2- reestructuración. La razón de estos cambios es lograr un análisis de datos congruente y certero.

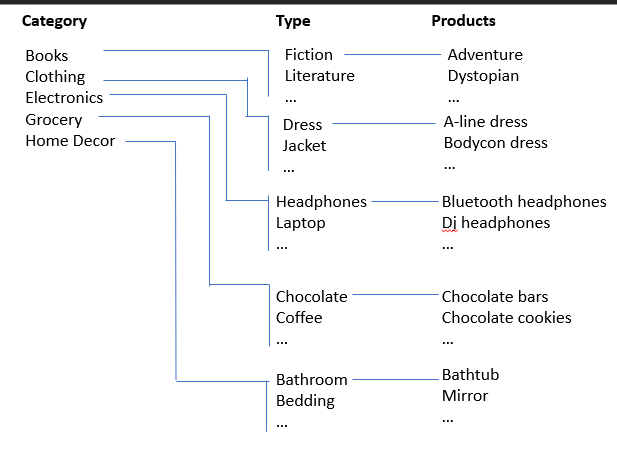
# Correcciones a la BD

1. Se eliminan los atributos “Name”, “Email”, “Phone”, “Address”, “Zipcode”, “Shipping\_Method”, “Year”, “Month” y “Total \_Amount”.
2. En el atributo “Product\_brand” corregir cambiar los campos “Mitsubhisi” por “Mitsubishi” y “Whirepool” por “Whirlpool”.
3. Los campos “BlueStar AC” y “Mitsubishi 1.5 Ton 3 Star Split AC”, pertenecientes a el atributo “Product\_Type”, ambos se cambian al nuevo campo “Air Conditioner”.
4. Rellenar el atributo “Product\_Category” según su “Product\_Type” correspondiente. Por ejemplo, en las primeras dos filas sabemos que el tipo “Bathroom” tiene como categoría “Home Decor”.

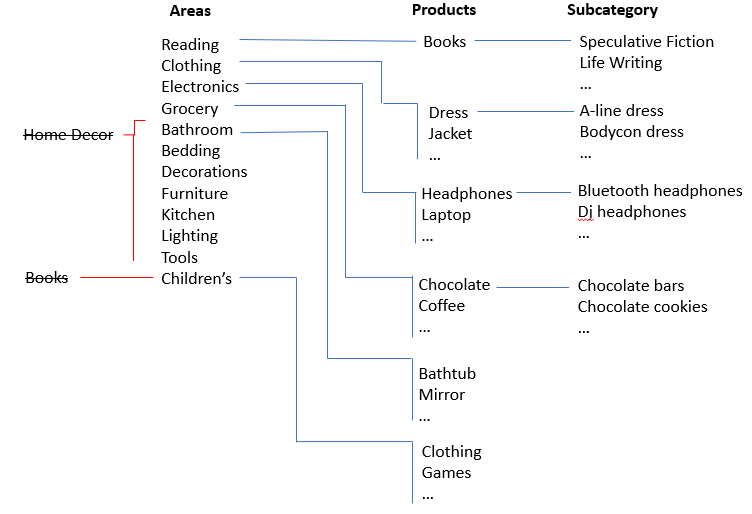
|  |  |
| --- | --- |
| product\_category | Product\_Type |
| NULL | Bathroom |
| Home Decor | Bathroom |
| NULL | Bedding |
| Home Decor | Bedding |
| Electronics | BlueStar AC |
| NULL | Children's |
| Books | Children's |
| NULL | Chocolate |
| Grocery | Chocolate |
| NULL | Coffee |
| Grocery | Coffee |
| NULL | Decorations |
| Home Decor | Decorations |
| NULL | Dress |
| Clothing | Dress |
| NULL | Fiction |
| Books | Fiction |
| Electronics | Fridge |
| NULL | Furniture |
| Home Decor | Furniture |
| NULL | Headphones |
| Electronics | Headphones |
| NULL | Jacket |
| Clothing | Jacket |
| NULL | Jeans |
| Clothing | Jeans |
| NULL | Juice |
| Grocery | Juice |
| NULL | Kitchen |
| Home Decor | Kitchen |
| NULL | Laptop |
| Electronics | Laptop |
| NULL | Lighting |
| Home Decor | Lighting |
| NULL | Literature |
| Books | Literature |
| Electronics | Mitsubishi 1.5 Ton 3 Star Split AC |
| NULL | Non-Fiction |
| Books | Non-Fiction |
| NULL | Shirt |
| Clothing | Shirt |
| NULL | Shoes |
| Clothing | Shoes |
| NULL | Shorts |
| Clothing | Shorts |
| NULL | Smartphone |
| Electronics | Smartphone |
| NULL | Snacks |
| Grocery | Snacks |
| NULL | Soft Drink |
| Grocery | Soft Drink |
| NULL | T-shirt |
| Clothing | T-shirt |
| NULL | Tablet |
| Electronics | Tablet |
| NULL | Television |
| Electronics | Television |
| NULL | Thriller |
| Books | Thriller |
| NULL | Tools |
| Home Decor | Tools |
| NULL | Water |
| Grocery | Water |

# Reestructuración de la BD

Para lograr la comparación congruente entre los distintos campos de la BD se buscó reestructurar la base de datos. En la siguiente imagen se puede ver un diagrama con las relaciones originales de la BD.



En la siguiente imagen se puede ver el nuevo esquema.



1. Las columnas “Product\_Category” y algunos “Product\_Type” se eliminan. Se combinan sus contenidos en una nueva columna llamada “Areas”:
   1. Las categorías “Clothing”, “Electronics” y “Grocery” pasan a estar en “Areas”
   2. Se crea una nueva área llamada “Reading”.
   3. La columna “Home Decor” se elimina. Todos los tipos dentro de la categoría “Home Decor” pasan a estar dentro de “Areas”.
   4. El tipo “Children’s” de la categoría “Books” pasa a estar dentro de “Areas”
2. La columna “Products” se renueva desde cero:
   1. Los tipos originalmente pertenecientes a “Clothing”, “Electronics” y “Grocery” pasan a estar en la columna “products”.
   2. Toda la categoría “Books” pasa a estar dentro de “Products”.
   3. Los “Products de “Home Decor” se mantienen al igual que los productos del tipo “Children’s” exceptuando el producto “Books”.
3. Se agrega una nueva columna llamada “Subcategory” y se colocan los productos más específicos:
   1. Todos los productos de “Clothing” y “Grocery” pasan a estar dentro de “Subcategory”. Se hace lo mismo para los productos de “Electronics” exceptuando los que tengan marca “Apple”, “Samsung” o “Sony”, También se exceptúan aquellos que no tengan marca (valores NULL).
   2. Ni “Home Decor” ni “Children’s” tienen campos dentro de “Subcategory”.
   3. Se crea una nueva clasificación para los “Products” de la categoría original “Books”, es tal como sigue:
      1. Speculative Fiction: Fantasy, Science Fiction y Dystopian
      2. Realistic Fiction: Literary Fiction y Historical Fiction
      3. Genre Fiction: Adventure Horror, Mystery, Romance y Action
      4. Life Writing: Biography y Memoir
      5. Professional & Academic: Business, Psychology y Science
      6. Practical & Lifestyle: Cooking, Health, Self-help y Travel
      7. Thriller: Crime, Detective, Legal Thriller, Political Thriller, Techno-thriller, Psychological Thriller y Suspense
      8. Classical/Canonical: Classic Literature, Drama y Poetry
      9. Modern/Contemporary: Contemporary Literature, Modern Literature y Short Stories
      10. Forms & Styles: Anthologies y Essays