**Web scraping project - SHEIN cloths website ERD**

Authors: Limor Nunu & Dana Makov

-------------------------------------------------------------------------------

Overview:

After the user ask to scrap products from the site, the data saved in the data base in few tables: one main table and secondary tables.

The main table contains key, name of the product type and few main data fields collected from the website as price and rating.

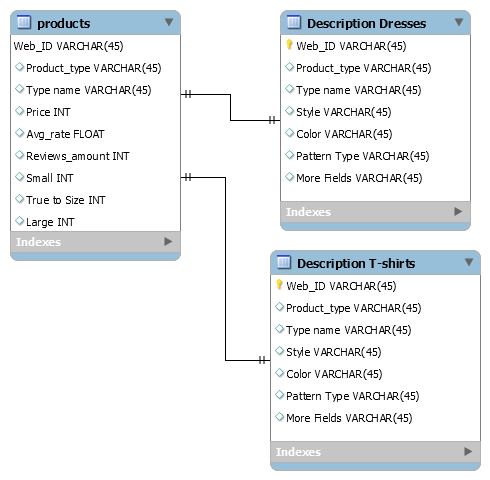
The secondary tables are products tables divided by product type.

In each secondary table there is key and more data fields (description) of the product, collected from the website as color and style.

The entities we have here are the products and the relation between them is “one-to-one”, because every item has its own description.

The primary key of the main table is the ID of the product taken from the website; it’s also the foreign key in the tables by product type.

Let’s demonstrate this in the next diagram:



The Web ID connects between the tables.

The reason why we construct this in that way:

If we look on the data fields for each product we’ll see that for other group of products there are different description fields (e.g. sleeves are relevant for dresses and t-shirts but not for swimwear).

Therefore, we decided to split the descriptive characters of the products to different tables.

As we mentioned before, the main table that called “products” contains all the products items that were scrapped by ID, product type and few main data fields as price and average rating and the secondary tables (named by the product type as dresses) will contain other descriptive data.

The columns shown in each table:

|  |  |
| --- | --- |
| * *Primary\_key* | Web\_ID |
|  | Product\_type |
|  | Price |
|  | Average\_rating |
|  | Reviews\_amount |
|  | Small |
|  | True\_to\_Size |
|  | Large |

**Products:**

**Dresses Description:**

|  |  |
| --- | --- |
| * *Primary\_key* | Web\_ID |
|  | Product\_type |
|  | Style |
|  | Color |
|  | Pattern Type |
|  | Neckline |
|  | Dresses Length |
|  | Type |
|  | Details |
|  | Sleeve Length |
|  | Season |
|  | Composition |
|  | Material |
|  | Fabric |
|  | Waist Line |
|  | Sheer |
|  | Hem Shaped |
|  | Fit Type |
|  | Belt |
|  | Sleeve Type |
|  | Lining |

**T-shirts Description:**

|  |  |
| --- | --- |
| * *Primary\_key* | Web\_ID |
|  | Product\_type |
|  | Style |
|  | Color |
|  | Pattern Type |
|  | Neckline |
|  | Length |
|  | Type |
|  | Details |
|  | Season |
|  | Composition |
|  | Material |
|  | Fabric |
|  | Sheer |
|  | Fit Type |
|  | Sleeve Length |
|  | Sleeve Type |
|  | Placket Type |
|  | Arabian Clothing |

**Swimwear Description:**

|  |  |
| --- | --- |
| * *Primary\_key* | Web\_ID |
|  | Product\_type |
|  | Style |
|  | Color |
|  | Pattern Type |
|  | Neckline |
|  | Composition |
|  | Material |
|  | Fabric |
|  | Bra Type |
|  | Bottom Type |
|  | Lining |
|  | Chest pad |