Data dictionary Wehkamp database

The data will be provided to you in a PostgreSQL database, that runs on the computer systems of the University of Groningen.

In the database, you find the following tables:

Article table

Article is a table with dimensional data about the articles sold by Wehkamp, limited to category “Beachwear”.

Column	**Description**
ARTICLE_ID	Identification for the article
CLASS_1	First-level merchandise classification, used internally.
CLASS_2	Second-level merchandise classification, used internally.
CLASS_3	Third-level merchandise classification, used internally.
BRAND_NAME	Identification of the brand name
LIVE_YEAR	The year the article was live on platforms

Customers table

Customers is a table with dimensional data regarding customers of Wehkamp. The set is a sample of all customers of Wehkamp.

Column	**Description**
CUSTOMER_ID	Identification of the customer
GENDER_CODE	Identification of the gender
START_DATE_YEAR	The year the person became customer
GEOM_URBANISATION	Degree of urbanisation: 1.00 – High urbanisation 2.00 3.00 4.00 5.00 – Low urbanisation
GEOM_HOUSEHOLD_COMPOSITION	Composition of the household: 0. Unknown 1. Young singles 2. Middle-aged singles 3. Older singles 4. Families with only young children 5. Families with older children 6. Young couples without children 7. Middle-aged couples without children 8. Older couples without children

GEOM_HOUSEHOLD_AGE	Age category of the household: 0. Unknown 1. < 25 years 2. 25 - 29 years 3. 30 - 34 years 4. 35 - 39 years 5. 40 - 44 years 6. 45 - 49 years 7. 50 - 54 years 8. 55 - 59 years 9. 60 - 64 years 10. 65 - 69 years 11. 70 - 74 years 12. 75 - 79 years 13. >= 80 years
GEOM_HOUSEHOLD_INCOME	Income category of the household: 0. Unknown 1. < 18,000 2. 18,000 - 26,000 3. 26,000 - 35,000 4. 35,000 - 50,000 5. 50,000 - 75,000 6. 75,000 - 100,000 7. 100,000 - 200,000 8. >= 200,000
GEOM_EDUCATION	Education level of the household: 0. Unknown 1. Primary education 2. LBO/VMBO-K/VMBO-B/MBO 1 3. MAVO/MULO/VMBO-T/VMBO-G 4. MBO (2, 3 or 4) 5. HAVO/VWO/HBS 6. HBO or WO bachelor 7. HBO or WO master/MBA/postdoc
GEOM_READS_NEWSPAPER	Whether the household reads newspapers: 0. Unknown 1. No 2. Yes
GEOM_PREFERRED_NEWSPAPER_TYPE	Preferred newspaper type: 0. Unknown

	1. No newspaper 2. Physical newspaper 3. No preference 4. Digital newspaper
GEOM_CONSUMPTION_FREQUENCY	Consumption frequency of the household: 0. Unknown 1. Little 2. Average 3. Much
GEOM_CLOTHING_SEGMENT	The type of clothing segment best applicable to the household: 0. Unknown 1. Budget 2. Fashionable 3. Middle-class 4. Exclusive 5. Second hand
GEOM_CLOTHING_BUDGET	Budget allocated to clothing: 0. Unknown 1. Little 2. Below average 3. Average 4. Above average 5. Much
GEOM_CLOTHING_FASHIONABLE	Estimation how much cloths are being bought in the ‘fashionable’ segment 0. Unknown 1. Little 2. Below average 3. Average 4. Above average 5. Much

Events table

Events is a table in which every line represents an event of a visitor regarding an article (view/cart/buy).

Remarks:

- Usually, the conversion for a given article/date range is calculated by dividing the number of sales (purchase) by the number views (product detail view) in that range.
- Not all events are registered. For instance, article impressions (e.g. product overview pages) also show articles, but these impressions are not in the set.
- A customer can have multiple events per session.

Column	**Description**
ARTICLE_ID	Identification of the article
INTERNET_SESSION_ID	Identification for the session
ACTION_TYPE_DESC	What did the visitor do with the article? Possible values: ● add to cart ● product detail view ● purchase ● remove from cart
INTERNET_SESSION_DTIME	Date/time of article events
CUSTOMER_ID	Identification of the customer
DEVICE_CATEGORY_DESC	Category of the device that was used during the event: ● desktop ● mobile ● tablet
SESSION_DURATION	Length of the session

Orders table

Orders table contains order factual data. An order can consist of one or more order lines. The order dataset contains one record for each order line.

Column	**Description**
DATE_DATE	The date of the order
ORDER_ID	Identification of the order
ORDERLINE_ID	Identification of the order line
ARTICLE_ID	Identification of the ordered article
CUSTOMER_DEMAND_AMT	Demand amount of the item sold
RETURN_QTY	Quantity of items returned. If the item is returned, the return quantity will be 1
CUSTOMER_ID	Identification of the customer
ARTICLE_SKU	Second level identification of the article. This can be because of a different color or product size.