DATA MINING CUP 2021

items.csv

Column name	Description	Value Range / Format
itemID	Unique identifier for every product	Positive Integer
title	Title of the product	String
author	Name of the author	String
publisher	Name of the publishing company	String
main topic	*	String
subtopics		List of Strings, surrounded by "[" and "]" topics are separated by "," and without space characters

* The main- and sub-topics come in the form of cryptic strings, where each character, or a set of characters if the set is divided by "-", denotes a category. Each topic string follows the path of a category hierarchy. With the leftmost character being the top level and the rightmost being the bottom level category. The topic strings can vary in length and do not necessarily follow the category hierarchy to its deepest possible level. Every product can have one main topic and a list of subtopics. The topic strings follow a general thematic subject category hierarchy, which is deciphered and explained at https://ns.editeur.org/thema/en.

transaction.csv

Column name	Description	Value Range / Format
itemID	Unique identifier for every product	Positive Integer
sessionID	Unique identifier for a transaction session	Positive Integer
click	Number of clicks for a product in a session	Integer
basket	Quantity of the product that is added into the basket during a session	Integer
order	Number of orders for a product in a session	Integer