

DATA MINING CUP 2022

items.csv

Column name	Description	Value Range / Format
itemID	Unique identifier for every product	Positive Integer [0, ..., 32775]
brand	Brand of the product	Integer* [-1, ..., 1513]
feature_1	Descriptive feature (categorical)	Integer* [-1, ..., 10]
feature_2	Descriptive feature (categorical)	Positive Integer {0, 1, 2, 3}
feature_3	Descriptive feature (categorical)	Integer* [-1, ..., 538]
feature_4	Descriptive feature (categorical)	Integer* {-1, 0, 1, 2, 3, 4}
feature_5	Descriptive feature (categorical)	Integer* [-1, ..., 190]
categories	List of associated categories	Comma-separated list of integers

* The value of -1 corresponds to no feature value.

category_hierarchy.csv

Column name	Description	Value Range / Format
category	Category identifier	Positive Integer [0, ..., 4299]
parent_category	Broader (parent) category associated with a category identifier	Positive Integer subset of the integer range of the column 'category'

orders.csv

Column name	Description	Value Range / Format
date	Date of the transaction	YYYY-MM-dd from 2020-06-01 to 2021-01-31
userID	Unique identifier for users	Positive Integer [0, ..., 46137]
itemID	Item identifier of the sold item	Positive Integer [0, ..., 32775]
order	Number of orders of a product at a given date by a specific user	Positive Integer

submission.csv

Column name	Description	Value Range / Format
userID	Unique identifier for users	Positive Integer, subset of the userIDs in orders.csv
itemID	Item identifier of the sold item	Positive Integer, subset of the itemIDs in items.csv
prediction	Column to be filled by the participating teams	Positive Integer {0, 1, 2, 3, 4}