

**DATA:** You are provided with a year of transactional data for 22,873 users on the site ponpare.jp. The training set spans the dates 2011-07-01 to 2012-06-23. The test set spans the week after the end of the training set, 2012-06-24 to 2012-06-30. The goal of the competition is to recommend a ranked list of coupons for each user in the dataset (found in *user\_list.csv*). Your predictions are scored against the actual coupon purchases, made during the test set week, of the 310 possible test set coupons.

File descriptions

The dataset has relational format, with hashed ID columns for each entity.

- **user\_list.csv** - the master list of users in the dataset

Column Name	Description	Type	Length	Decimal	Note
USER_ID_hash	User ID	VARCHAR2	32		
REG_DATE	Registered date	DATE			Sign up date
SEX_ID	Gender	CHAR	1		f = female m = male
AGE	Age	NUMBER	4	0	
WITHDRAW_DATE	Unregistered date	DATE			
PREF_NAME	Residential Prefecture	VARCHAR2	2		[JPN] Not registered if empty

- **coupon\_list\_train.csv** - the master list of coupons which are considered part of the training set
- **coupon\_list\_test.csv** - the master list of coupons which are considered part of the test set. Your competition predictions should be sourced only from these 310 coupons. You will not receive credit for predicting training set coupons that were purchased during the test set period.

Column Name	Description	Type	Length	Decimal	Note
CAPSULE_TEXT	Capsule text	VARCHAR2	20		[JPN]
GENRE_NAME	Category name	VARCHAR2	50		[JPN]
PRICE_RATE	Discount rate	NUMBER	4	0	
CATALOG_PRICE	List price	NUMBER	10	0	
DISCOUNT_PRICE	Discount price	NUMBER	10	0	
DISPFROM	Sales release date	DATE			
DISPEND	Sales end date	DATE			
DISPPERIOD	Sales period (day)	NUMBER	4	0	
VALIDFROM	The term of validity starts	DATE			
VALIDEND	The term of validity ends	DATE			
VALIDPERIOD	Validity period (day)	NUMBER	4	0	
USABLE_DATE_MON	Is available on Monday	CHAR	1		
USABLE_DATE_TUE	Is available on Tuesday	CHAR	1		
USABLE_DATE_WED	Is available on Wednesday	CHAR	1		
USABLE_DATE_THU	Is available on Thursday	CHAR	1		
USABLE_DATE_FRI	Is available on Friday	CHAR	1		
USABLE_DATE_SAT	Is available on Saturday	CHAR	1		
USABLE_DATE_SUN	Is available on Sunday	CHAR	1		
USABLE_DATE_HOLIDAY	Is available on holiday	CHAR	1		
USABLE_DATE_BEFORE_HOLIDAY	Is available on the day before holiday	CHAR	1		

large_area_name	Large area name of shop location	VARCHAR2 30	[JPN]
ken_name	Prefecture name of shop	VARCHAR2 8	[JPN]
small_area_name	Small area name of shop location	VARCHAR2 30	[JPN]
COUPON_ID_hash	Coupon ID	VARCHAR2 32	

- **coupon\_visit\_train.csv** - the viewing log of users browsing coupons during the training set time period. You are not provided this table for the test set period.

Column Name	Description	Type	Length	Decimal	Note
PURCHASE_FLG	Purchased flag	NUMBER	1	0	0:Not purchased 1:Purchased
PURCHASEID_hash	Purchase ID	VARCHAR2	128		
I_DATE	View date	DATE			Purchase date if purchased
PAGE_SERIAL		VARCHAR2			
REFERRER_hash	Referer	VARCHAR2	4000		
VIEW_COUPON_ID_hash	Browsing Coupon ID	VARCHAR2	128		
USER_ID_hash	User ID	VARCHAR2	10		
SESSION_ID_hash	Session ID	VARCHAR2	128		

- **coupon\_detail\_train.csv** - the purchase log of users buying coupons during the training set time period. You are not provided this table for the test set period.

Column Name	Description	Type	Length	Decimal	Note
ITEM_COUNT	Purchased item count	NUMBER	10	0	
I_DATE	Purchase date	DATE			
SMALL_AREA_NAME	Small area name	VARCHAR2	30		[JPN] User residential area name
PURCHASEID_hash	Purchase ID	VARCHAR2	32		
USER_ID_hash	User ID	VARCHAR2	32		
COUPON_ID_hash	Coupon ID	VARCHAR2	32		

- **coupon\_area\_train.csv** - the coupon listing area for the training set coupons
- **coupon\_area\_test.csv** - the coupon listing area for the test set coupons

Column Name	Description	Type	Length	Decimal	Note
SMALL_AREA_NAME	Small area name	VARCHAR2	30		[JPN]
PREF_NAME	Listed prefecture name	VARCHAR2	8		[JPN]
COUPON_ID	Coupon ID	VARCHAR2	32		

- **sample\_submission.csv** - a sample file showing the correct format for predictions
- **documentation.zip** - an archive of Excel files containing an entity relationship diagram and English translations

