1. the goal and meaning of this data analysis
   we have some data for the product information and user information, and the user's action information with respect to the product during 2016-02-01 to 2016-04-15. My final goal was to use all this information to recommend the products to customers that they are more likely to be interested in.

2. the process of the whole project
   1> data analysis
      we need to do some statistical analysis on the user data, product data, and action data
   2> process for the NA and other things
   3> split the data into in-sample data and out-of-sample data
   4> use the in-sample data to train model and test it on the out-of-sample data

3.

Data description

User data

user_id	meaning
age	-1 to denote no info
sex	0:male，1:female，2:no info
user_lv_cd	User level . Higher number, higher level
user_reg_dt	User registration day

Product information

sku_id	explanation
attr1	Enum type. -1 to denote no info
attr2	Enum type. -1 to denote no info
attr3	Enum type. -1 to denote no info
cate	Class ID
brand	Brand ID

Comment information

dt	Comment deadline day
sku_id	Id of product
comment_num	The number of comments for that product 0: no comment 1: 1 comment 2: 2-10 comments 3: 11-50 comments 4: more than 50 comments

has_bad_comment	0: have no bad comment, 1:have bad comment
bad_comment_rate	The ratio of bad comments to totoal comments

Action information

user_id	The id of product's user
sku_id	The id of product
time	The time of action
model_id	The module that user click if click on the product
type	The type of the action: 1. Browse 2. Add to the shopping list 3. Delete from the shopping list 4. Purchase 5. Add the product to the interest list 6. click
cate	Class ID
brand	Brand ID