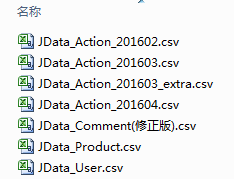
**JData算法大赛FAQ**

**赛题篇**

1、 Q：赛题数据量级

A：本次赛题数据共7个csv文件。



2、 Q：赛题数据问题件的编码格式

A：提供的数据文件编码格式为GBK。

3、 Q：评测数据中一个用户是否存在购买多个商品？

A：不存在，一个用户（user\_id）只购买一个所提供的候选商品集合中的商品（sku\_id）；用户购买候选商品集合之外的商品无需提交。

4、 Q：关于商品集合P和S

A：集合P为候选商品集合，即参赛者预测的结果中的sku需要在集合P中；集合S为行为数据中的商品全集；

5、 Q：评论数据是当天的还是累计的

A：评论数据为截止到当日的累计数据

6、 Q：关于model\_id的解释

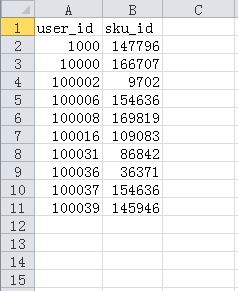
A：model\_id表明用户在页面上点击了哪一个位置；数据中可能存在一些空值，有可能是异常值，也有可能是在页面上点击了一个空白的位置产生的数据，请参赛者自行理解并处理

7、 Q：提交数据文件有什么要求？

A： 1）提交数据文件需要去重处理，即：提交结果文件中用户user\_id唯一，不存在重复；

2）预测用户下单的商品sku\_id，必须在商品子集P中；

3）提交数据文件格式为csv，编码为utf-8，第一行为表头，即：第一列为“user\_id”，第二列为“sku\_id”，提交数据文件截图如下：



8、 Q：关于日期和时间格式

A：日期格式统一为"yyyy-mm-dd"，时间格式统一为"yyyy-mm-dd hh:mi:ss"

9、 Q：关于评价数据中的有差评但是差评率为0的情况

A：这种情况是因为第一版数据中存在一些小问题，我们已经进行了紧急修正，如果是下载了第一版数据的同学，请重新下载修正版评价数据（压缩包文件名：JData\_Comment(修正版)）。

10、Q：关于3月12~3月20的行为数据缺失情况

A：目前数据已经补充完成，如果是下载了第一版数据的同学，请下载3月份行为补充数据（压缩包文件名：JData\_Action\_201603\_extra.csv）。

11、Q：关于用户表中存在10条记录，用户注册时间晚于2016年4月15日情况

A：此情况为真实生产环境中的异常数据，请选手自行理解并处理。

**奖励篇**

1、 Q：获奖参赛队伍每个成员是否都能拿到Special Offer？

A：奖项中所涉及的Special Offer均是真Special Offer，只要获奖参赛团队中成员通过专家评委的考核，就能拿到Special Offer，参赛团队成员通过几人，京东就敢发出几个Special Offer；获得冠亚季军的每个团队至少发出一个Special Offer。