

数据说明

本次比赛提供产品信息和销量相关数据，格式为UTF-8编码格式的txt文件，以逗号分隔。
产品信息表product_info.txt，4000行数据；价格与销量表product_quantity.txt，1959676行数据；提交样例prediction_lilei_20170320.txt，56000行数据。

备注：1) 提供的数据经过了脱敏，和产品的实际信息有一些差距，但是不会影响这个问题的可解性；2) -1代表缺失值。

product_info.txt：产品信息表

变量名	类型	说明	示例
product_id	int	产品id	3179
district_id1	int	一级地区id	10201
district_id2	int	二级地区id	20502
district_id3	int	三级地区id	31003
district_id4	int	四级地区id	45448
lat	double	纬度	3.994472504
lon	double	经度	11.6463295
railway	double	距火车站距离	8.1
airport	double	距机场距离	20.2
citycenter	double	距市中心距离	10.3
railway2	double	距第二火车站距离	0.9
airport2	double	距第二机场距离	1
citycenter2	double	距市副中心距离	1.1
eval	int	官方评级	6
eval2	int	携程评级	7
eval3	double	用户评级	5.5
eval4	double	综合评分	3.4
voters	int	点评人数	7431
startdate	string	开售时间	1984-01-01
upgradedate	string	升级时间	2007-01-01
cooperatedate	string	合作时间	2013-02-27
maxstock	int	最大库存数，即每日可售数量上限	910

product_quantity.txt: 价格与销量表

变量名	类型	说明	示例
product_id	int	产品id	3179
product_date	string	产品使用日期，格式为yyyy-mm-01，取值范围为2014-01-01至2015-11-30	2015-01-30
orderattribute1	int	订单属性1： 1， 2， 3	1
orderattribute2	int	订单属性2： 1， 2	2
orderattribute3	int	订单属性3： 1， 2	1
orderattribute4	int	订单属性4： 1， 2	1
ciiquantity	bigint	成交量	18
ordquantity	bigint	预订量	20
price	int	平均售价	572

prediction_lilei_20170320.txt: 提交样例

变量名	类型	说明	示例
product_id	int	产品id	3179
product_month	string	产品使用月份，格式为yyyy-mm-01，取值范围为2015-12-01至2017-01-01，即 <u>2015年12月至2017年1月共14个月</u>	2016-08-01
ciiquantity_month	bigint	月成交量	2200