

比赛网址：

https://www.kesci.com/apps/home_log/index.html#!/competition/58bfc27471db03332e1b8a36/content/0

背景介绍

携程作为中国领先的综合性旅行服务公司，每天向超过 2.5 亿会员提供全方位的旅行服务，海量的网站访问产生了海量的数据，从中挖掘潜在的信息资源有着重要意义。调研发现，在出行产品预订业务中，不同区域的产品需求量级不一样，不同时段需求量会有高低起伏，相同区域相同时段各产品的需求量因产品特性不同又有差异。

本数据集为携程脱敏后的产品信息和销量相关数据，希望通过对本数据集进行探索，深入了解产品需求量和产品特性、历史销量的关系，挖掘出影响需求量的关键因素，预测出行产品未来 14 个月每月的销量，从而指导产品的库存管理和定价策略

数据字典

1.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
2.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 |

数据分析

看 readme.html