## 网络零售平台异常商品（价格、销量）识别赛题简介

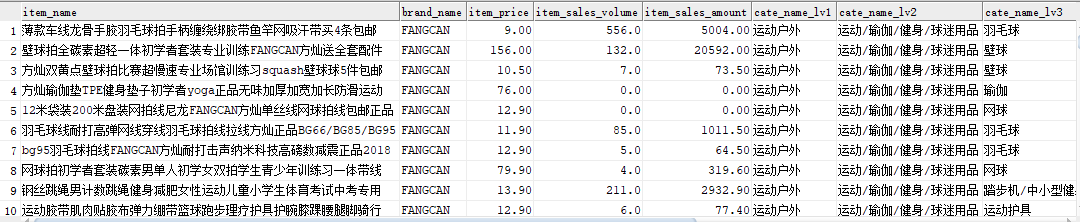
### 1.赛题数据

本赛题提供数据包含商品数据和商家数据两个部分：

（1）商品数据包含商品名称，商品品牌，商品所属类目，商品参数，商品价格，商品月销量，商品月销售额，评论数，库存，商品所属的店铺等基本信息。

（2）商家数据包含商家名称，商家当月经营商品数，商家当月商品销量，商家当月商品销售额，商家主营类型，商家所属省市区县，商家评分等店铺基本信息。

主要商品信息：（所有字段以及说明见附件1）





主要商家信息：（所有字段以及说明见附件2）



### 2.赛题目标

“网购”因其方便快捷、省时省力、送货上门的特点越来越受到人们的青睐。在各平台规模不断扩大、商品数不断增加的同时，一些不正当的经营行为，例如虚标价格、刷单行为也随之出现，严重违反了电商法，需要对这类商品数据进行准确识别。针对如此庞大的商品数量，如果单纯人工检查筛选，不仅工作量巨大，还会出现遗漏和错误的情况。本赛题旨在寻找一种识别方法，能够实现快速对异常商品的准确定位，减少人工干预成本和降低出错率。

### 3.赛题引导

本赛题发散性强，参赛选手可以根据提供的数据，发挥自己的想象力，探索不同的解决方案。

由于数据量巨大，且数据指标较多，数据相对复杂，很难通过建立一系列的固定规则或者规范对异常数据进行筛选。本赛题主要是希望参赛者，通过对数据的探索，发现或者发掘一种异常数据识别系统或者模型办法，能够对商品数据进行快速过滤，识别其中的异常。

部分商品价格异常数据具有一定的离散，可以考虑基于商品的数据维度（多维度分类分析等）进行离散分析。可以从网上查阅商品刷单的相关资料，来对商品销量异常进行识别。本赛题提供商品和店铺两个大类的数据，参赛者可以结合使用，可以分析商品的稳定性和店铺经营的稳定性，参考部分电商行业经营的业务知识。

异常商品价格举例：

（1）绿色植物多肉容易养活 9999元 月销量5000件

（2）摩卡咖啡xx品牌咨询客服9.9 39.9元 月销量250000件

（3）蜜蜡手串直播间出售 1000元 月销量100000件

**提示：通过以上价格异常示例，建立类似提取规则，无法完整获取提供的商品中的异常值**

异常商品销量举例：（提示品牌、销量、店铺、所属地区等都是分析维度）

（1）xxx保暖棉衣个性 A品牌 159元 月销量15001件 A旗舰店

（2）xxx保暖卫衣年轻 A品牌 139元 月销量12001件 A旗舰店

**提示：通过以上销量异常示例，建立类似提取规则，无法完整获取提供的商品中的异常值**

### 4.赛题要求

（1）建立的识别系统或者模型方法，能够识别出异常商品（模型方法识别出的异常商品占总异常商品数的比例越大越好），但是识别精度要高（模型方法识别出的异常值准确率）。

（2）建立的识别系统或者模型方法，识别效率高（考虑时间复杂度和空间复杂度）。

附录：

1.商品信息表字段以及说明

|  |  |  |
| --- | --- | --- |
| 序号 | 字段 | 字段说明 |
| 1 | ITEM\_ID | 商品id |
| 2 | ITEM\_NAME | 商品名称 |
| 3 | BRAND\_ID | 品牌id |
| 4 | BRAND\_NAME | 品牌名称 |
| 5 | ITEM\_PRICE | 商品价格 |
| 6 | ITEM\_ORIGINAL\_PRICE | 商品划线价 |
| 7 | ITEM\_SALES\_VOLUME | 商品月销量 |
| 8 | ITEM\_SALES\_AMOUNT | 商品月销售额 |
| 9 | ITEM\_URL | 商品url |
| 10 | CATE\_ID\_LV1 | 一级类目id |
| 11 | CATE\_NAME\_LV1 | 一级类目名称 |
| 12 | CATE\_ID\_LV2 | 二级类目id |
| 13 | CATE\_NAME\_LV2 | 二级类目名称 |
| 14 | CATE\_ID\_LV3 | 三级类目id |
| 15 | CATE\_NAME\_LV3 | 三级类目名称 |
| 16 | CATE\_ID\_LV4 | 四级类目id |
| 17 | CATE\_NAME\_LV4 | 四级类目名称 |
| 18 | CATE\_ID\_LV5 | 五级类目id |
| 19 | CATE\_NAME\_LV5 | 五级类目名称 |
| 20 | TOTAL\_EVAL\_NUM | 评论数 |
| 21 | ITEM\_FAV\_NUM | 收藏数 |
| 22 | ITEM\_STOCK | 库存 |
| 23 | ITEM\_DELIVERY\_PLACE | 发货地 |
| 24 | ITEM\_PROD\_PLACE | 产地 |
| 25 | ITEM\_PARAM | 商品参数 |
| 26 | USER\_ID | 店铺ID |
| 27 | SHOP\_NAME | 店铺名 |
| 28 | SHOP\_PROVINCE | 店铺省份 |
| 29 | SHOP\_CITY | 店铺城市 |
| 30 | SHOP\_OPEN\_DATE | 开店时间 |
| 31 | SHOP\_OPEN\_YEARS | 开店时长 |
| 32 | COMPANY\_PROVINCE | 公司省份 |
| 33 | COMPANY\_CITY | 公司城市 |
| 34 | COMPANY\_COUNTY | 公司区县 |
| 35 | SHOP\_DELIVERY\_PROVINCE | 主营发货省份 |
| 36 | SHOP\_DELIVERY\_CITY | 主营发货城市 |
| 37 | SHOP\_DELIVERY\_COUNTY | 主营发货区县 |

2.商家信息表字段以及说明

|  |  |  |
| --- | --- | --- |
| 序号 | 字段 | 字段说明 |
| 1 | USER\_ID | 店铺ID |
| 2 | SHOP\_NAME | 店铺名 |
| 3 | COMPANY\_NAME | 公司名称 |
| 4 | SHOP\_SALES\_VOLUME | 店铺月销量 |
| 5 | SHOP\_SALES\_AMOUNT | 店铺月销售额 |
| 6 | MAIN\_BUSINESS | 店铺主营类型（基于平台类目） |
| 7 | BUSINESS\_SCOPE | 店铺经营范围 |
| 8 | SHOP\_PROVINCE | 店铺省份 |
| 9 | SHOP\_CITY | 店铺城市 |
| 10 | SHOP\_OPEN\_DATE | 开店时间 |
| 11 | SHOP\_OPEN\_YEARS | 开店时长 |
| 12 | SHOP\_URL | 店铺URL |
| 13 | ITEMDESC\_SCORE | 店铺商品描述满意度 |
| 14 | SERVICE\_SCORE | 店铺服务满意度 |
| 15 | DELIVERY\_SCORE | 店铺物流满意度 |
| 16 | ITEMDESC\_RANK | 店铺商品描述满意度排名 |
| 17 | SERVICE\_RANK | 店铺服务满意度排名 |
| 18 | DELIVERY\_RANK | 店铺物流满意度排名 |
| 19 | SHOP\_LOCATE | 店铺地址 |
| 20 | COMPANY\_PROVINCE | 公司省份 |
| 21 | COMPANY\_CITY | 公司城市 |
| 22 | COMPANY\_COUNTY | 公司区县 |
| 23 | SHOP\_DELIVERY\_PROVINCE | 主营发货省份 |
| 24 | SHOP\_DELIVERY\_CITY | 主营发货城市 |
| 25 | SHOP\_DELIVERY\_COUNTY | 主营发货区县 |