**数据说明与提交规则**

## 赛题：线上线下全场景生鲜超市库存履约一体化决策

## 训练数据

### 商品基础信息表 sku\_info.csv

主键：sku\_id

|  |  |  |
| --- | --- | --- |
| 列名 | 数据样例 | 说明 |
| sku\_id | 5 | 商品编号（唯一标识） ，一个商品编号对应唯一商品种类； |
| item\_first\_cate\_cd | 2 | 商品一级类目编号，枚举值 |
| item\_second\_cate\_cd | 6 | 商品二级类目编号，枚举值 |
| item\_third\_cate\_cd | 8 | 商品三级类目编号，枚举值 |
| brand\_code | 1 | 商品品牌编号，枚举值 |

### 商品销售表 sku\_sales.csv

主键：order\_id + store\_id + sku\_id

|  |  |  |
| --- | --- | --- |
| 列名 | 数据样例 | 说明 |
| order\_id | 22 | 订单编号（订单唯一标识） |
| store\_id | 10 | 门店编号 |
| sku\_id | 5 | 商品编号 |
| order\_time | 2022-08-16 01:03:55 | 订单日期+时间，时间仅可用于排序，非真实时间，不影响问题求解 |
| quantity | 1 | 订单所购商品的数量 |
| channel | 1 | 下单渠道，；  枚举值，1：线下，2：线上 |

说明：

* 1. 字段quantity对应赛题中的；此表仅包含训练数据；
  2. 字段channel用于判断订单来源于线上还是线下；

### 商品价格和状态表 sku\_price\_and\_status.csv

主键：store\_id+sku\_id+date

|  |  |  |
| --- | --- | --- |
| 列名 | 数据样例 | 说明 |
| store\_id | 10 | 门店编号 |
| sku\_id | 5 | 商品编号 |
| date | 2023-02-13 | 销售日期 |
| salable\_status | 1 | 商品销售状态，  枚举值，0:下架不可售，1:上架可售 |
| stock\_status | 1 | 营业结束后报损前的商品剩余库存，评测日期范围内无意义；  枚举值，0：无剩余，1：有剩余 |
| original\_price | 15.08 | 商品原价 |

说明：

1. 此表包含训练和评测时间范围内的数据；记录从过去两年内商品在门店首次销售日起，对于从未在门店销售过的商品在该表没有记录；
2. 评测日期对应的salable\_status和stock\_status无参考意义，值为-1；
3. stock\_status为0时，说明当天商品可能缺货，消费者需求可能没完全释放，实际需求可能高于当日商品的实际订单销量总和；

### 商品促销信息表sku\_prom.csv

主键：store\_id + sku\_id + date + channel +promotion\_id，

若该表与其他表进行关联时，一定要注意该表的主键，store\_id + sku\_id + date + channel 有可能有重复记录，非唯一；

|  |  |  |
| --- | --- | --- |
| 列名 | 数据样例 | 说明 |
| store\_id | 2 | 门店编号 |
| sku\_id | 100 | 商品编号 |
| date | 2022-02-13 | 销售日期 |
| channel | 1 | 促销渠道，只有通过对应渠道下单才可享受到该促销，枚举值，1：线下，2：线上 |
| promotion\_id | 97 | 促销编号，同一个促销编号下通常会有多个参与促销的SKU |
| curr\_day | 1 | 今天是该促销开始后的第几天 |
| total\_days | 2 | 该促销一共持续多少天 |
| promotion\_type | 1 | 促销类型编号，枚举值，（比如单品打折，买一赠一、满额打折等） |
| discount\_off | 0.3 | 促销折扣率, 数值越大折扣越大；  若满足促销使用条件时（见\*1中的举例说明），商品单位售价 ，否则 |
| threshold | 2.2 | 判断一个订单中的商品能否享受该促销的条件阈值（订单金额需要大于等于至少多少倍该商品原价，见\*（2）说明）；该字段代表了一个sku\_id的折扣价格能否生效的难度，该值越大，促销越难生效； |

说明：

1. \*促销条件是否满足举例说明：（用于判断促销何时生效，参赛选手非必须了解和计算）

（1）某个订单中必须包含商品促销信息表中记录的商品信息，使用store\_id, sku\_id, date, channel四个字段来判断。

（2）某个订单的商品总金额必须大于等于促销商品的阈值条件；

举例说明，假设有如下：

商品销售信息（来自于sku\_sales.csv）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| order\_id | **store\_id** | **sku\_id** | **order\_time** | quantity | **channel** |
| 20230801001 | 1 | 1 | 2023-08-01 01:01:01 | 2 | 1 |
| 20230801001 | 1 | 2 | 2023-08-01 01:01:01 | 1 | 1 |

商品价格和状态信息（来自于sku\_price\_and\_status.csv）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **store\_id** | **sku\_id** | **date** | salable\_status | stock\_status | original\_price |
| 1 | 1 | 2023-08-01 | 1 | 1 | 25.88 |
| 1 | 2 | 2023-08-01 | 1 | 1 | 18.08 |

商品促销信息（来自于sku\_prom.csv）

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **store\_id** | **sku\_id** | **date** | **channel** | promotion\_id | curr\_day | total\_days | promotion\_type | discount\_off | threshold |
| 1 | 1 | 2023-08-01 | 1 | 123 | 1 | 2 | 1 | 0.3 | 2.2 |

通过表中标红的字段关联商品销售信息表、商品价格和状态信息表、商品促销信息表，我们可以统计到订单20230801001的优惠前订单总金额为 25.88 \* 2 + 18.08 \* 1 = 69.84元，同时商品1在促销表中有促销信息，阈值threshold是2.2倍商品原价，由于69.84 > 2.2 \* 25.88，因此订单20230801001所购的商品1可以享受促销优惠，即结算时的单位售价为25.88 \* （1 – 0.3）= 18.12元/件；而商品2没有促销信息，因此商品2的单位售价为商品原价18.08.

1. 此表包含训练和评测时间范围内的数据，此表只记录商品有促销时的促销信息，对于当天没有促销的商品，则不予记录；
2. 表中的样例数据解释如下：2022-02-13，商品100在门店2的线下销售渠道，有一条促销信息，促销编号是97，该促销共持续2天，2022-02-13是该促销开始的第一天，促销对应的类型是1；当用户在线下渠道购买商品100时，若优惠前订单金额大于等于商品100的商品原价的2.2倍时，用户订单中的商品100可以享受7折优惠（1 – 0.3），订单中的其他商品也可根据同样的逻辑来判断是否享受促销优惠；
3. 同一家门店下的同一个sku，在同一个销售渠道，同一天可能会有多个有效的促销(id)，分别对应着不同的适用条件（threshold），结算时用户的最终商品单位售价需根据当前用户下单渠道和商品购买实际情况来决定适用哪条促销规则；当同时满足多条促销规则时，生效的是优惠力度最大的促销规则。比如，某日某门店下的某商品在线上销售渠道同时具有购物满100打9折，满200打8折促销时，参赛选手需要根据综合情况，自行判断用户可能实际享受到的促销是9折还是8折。

### 门店区域天气表store\_weather.csv

主键：store\_id + date

|  |  |  |
| --- | --- | --- |
| 列名 | 数据样例 | 说明 |
| store\_id | 2 | 门店编号 |
| date | 2021-12-07 | 销售日期 |
| weather\_type | 1 | 天气类型编号，枚举值 |
| min\_temperature | 0.2 | 最低气温，归一化数值 |
| max\_temperature | 0.6 | 最高气温，归一化数值 |

1. 此表包含训练和评测时间范围内的数据；

### 评测程序中的关键字段计算方法

**这些字段用于说明评测时如何计算单位售价和商品成本价，参赛选手可以根据需要决定是否自行计算该些字段作为特征使用或者自行评测时使用。**

1. 未来评测期的单位售价，将根据未来的实际评测订单和未来促销信息，参照“商品促销信息表中的促销条件是否满足举例说明”，进行同理计算。即商品的单位售价（赛题中的），若未满足促销使用条件，单位售价=商品原价；若满足促销使用条件，单位售价 = 商品原价 \* （1 - discount\_off）；
2. 商品成本价（赛题中的），本赛题使用相对宽松条件，统一使用50%进行测评，即成本价=近7日平均商品售价（含当日） \* 0.5；

## 测评数据

### 评测目标表test\_item.csv

参赛者需要为表中的每一行记录提交前后场备货数量。

主键：store\_id + sku\_id + date

|  |  |  |
| --- | --- | --- |
| 列名 | 数据样例 | 说明 |
| store\_id | 2 | 门店编号 |
| sku\_id | 6 | 商品编号 |
| date | 2023-09-07 | 销售日期 |

* + store\_id范围： 1 ~ 12，共12家门店；
  + sku\_id范围： 1 ~ 1000，共1000种商品；
  + date范围： 2023-09-01 ~ 2023-09-14，共14天。

### 评测结果表test\_result.csv

参赛者需要按照评测结果表规定的格式和内容，为评测目标表中的全部评测对象提交参赛结果。

* + **文件格式**：标准csv文件，文件大小不超过10MB，英文逗号分隔，采用无BOM的utf8编码；
  + **文件内容**：全部门店的全部商品在未来日期范围内，每天的前场备货数量和后场备货数量，即使备货数量为0也需要包含；
  + **文件行数：**门店数 \* 商品数 \* 备货天数 + 1（header）。12\*14\*1000 + 1= 168001；
  + **文件样式**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| date | store\_id | sku\_id | x\_k | x\_m |
| 2023-09-01 | 1 | 2 | 231 | 250 |
| 2023-09-02 | 1 | 2 | 0 | 0 |
| 2023-09-14 | 5 | 66 | 0 | 5 |
| … | … | … |  |  |

共5列，分别为date, store\_id, sku\_id, x\_k, x\_m

其中：

date列表示备货日期，格式为yyyy-mm-dd，如2023-09-10；

store\_id列表示门店编号，正整数；

sku\_id列表示商品编号，正整数；

x\_k列表示前场备货数量，最多保留2位小数；

x\_m表示后场备货数量，最多保留2位小数。

### 方案的全部源代码和说明文档

**参赛选手需在规定的时间内，提交算法方案的全部代码和说明文档。**

**提交的格式**：zip包；

**提交的内容**：方案的全部源代码、程序运行环境说明、其他必要说明，外部辅助数据（若有的话）；方案需要支持可以复现提交的参赛结果。

## 选手离线自测流程

**参赛选手在迭代模型时，可以自行划定测试时间范围，构建测试集合、参考公开的评测程序进行离线效果自测。自测流程如下：**

### 准备测试商品范围及价格表 test\_sku\_prices.csv

为全部待测试商品在测试日期准备好价格数据，文件格式如下：

主键date + store\_id + sku\_id

|  |  |  |
| --- | --- | --- |
| 列名 | 数据样例 | 说明 |
| date | 2018-02-13 | 销售日期 |
| store\_id | 10 | 门店编号 |
| sku\_id | 5 | 商品编号 |
| salable\_status | 1 | 商品销售状态，  0:该行记录不参与测试，1:参与测试 |
| sale\_price | 20 | 商品单位售价，即赛题中的，计算方式见训练数据-商品促销信息表中的说明1 |
| cost\_price | 12 | 商品成本价，即赛题中的，计算方式见训练数据-商品促销信息表中的说明5 |
| ab\_type | A | 此处必须填’A’，否则该行记录不参与测试 |

### 准备测试集合对应的商品销售表 test\_sku\_sales.csv

根据选定的测试时间范围，把对应的销售订单按照商品销售表的格式准备好，具体格式见训练数据-商品销售表（sku\_sales.csv）；

### 准备评测结果表test\_result.csv

为全部待测试商品在测试日期准备好待验证的结果数据，具体格式见测评数据-测评结果表（test\_result.csv），注意test\_result.csv的文件行数需要和test\_sku\_prices.csv的文件行数一样。

### 执行评测

把准备好的test\_sku\_prices.csv、test\_sku\_sales.csv、test\_result.csv、与judge.py、test\_judge.py放入同一级目录下，执行test\_judge.py，观察自测结果。