Project executor: Jiale Chen

Project target description:

The data analysis group I served was in charge of contacting B side and communicating with Pin Duo Duo merchants. The objective of our project is to simplify the information rules of merchants' commodities. Typically, what we see on the C side are standard releasable commodity information after selection and the completion of risk control.

However, the commodity information prepared by merchants for the upload and release is most likely to be unqualified. It cannot be released on the platform as their information contains sensitive information, erotic image, political orientation, information from other platforms, etc.

There are two procedures for examining images, one of which is to utilize machines to examine the images uploaded for the commodities. When the machine fails, it will be turned over to a human for examination. However, there are some problems for the investigation by humans, including the inability to timely feedback the result of the examination, the duration of test being excessively long, the conditions being overly harsh for the rejection, the asymmetry of information, etc.

Therefore, our group proposed to move the module of machine examination from the position of submitting commodity information to the position of uploading commodity information, so that the merchants can learn if their commodity images and titles match the rules of the platform for the examination of commodities before the release of their commodities. However, there is a particular risk within this procedure as merchants of bad conduct may test the platform's rules that form an online confrontation. As soon as the merchants break the rules of bad conduct, they will likely skip the rules for the loophole that induces online badcase.

What the mentor expects me to do is that I need to do the data analysis on the risk and profit for the weight of advantages and disadvantages. And the analysis is mainly conducted by comparing indexes to check the overall situation and its influence on the merchant's experience.

I. Background and objectives:

1. Background of demand

Gauss:https://gauss.pdd.net/detail/85272

It is planned to move the examination rules for some commodities at postposition to preposition along with the management to improve the success rate of merchants for their release of commodities and their experience.

2. Objective of demand

Investigate into the current situation for the rejection of commodities on the platform and initially evaluate the project profit.

II. Assumption:

The rejection rate for the initial release in the whole channel is 10.75%, which is relatively high. Most merchants failed to release their commodities eventually after the rejection of the first release, which impacted their experience.

Both the proportion and the dynamic distribution rate of commodities released by new merchants through the APP for merchants are the highest for which they are of most value for the platform. However, the average submission with success after the rejection for the initial release is

about 6. We suggest putting it as a priority for the management.

III. Analytic process

1. Requirements and fact verification

- Data coverage
  - Date of statistics: November 1st 2021 to December 1st 2021
  - Merchant coverage: non Daipai merchants
  - Channel coverage: PC, open platform and merchants APP
  - Commodity coverage: newly submitted and released commodities (the creation time of the first examination corresponds to is_shop = 0)
- Index requirements:
  - Definition
    - Channel: channel corresponding to the creation time of first examination
    - New merchant: the date of the registered shop for the pass of first examination >= October 1st 2021
    - Dynamic distribution: the number of orders by the same day in history >0
    - Rejection to the first release: the creation time of the first examination corresponds to check_status=3
      - Success for the second time: the creation time of the second examination corresponds to check_status=2
      - Success in final: the creation time of the final examination corresponds to check_status=2
      - Failure in final: the examination time of the final examination corresponds to check_status IN (1, 3)
  - Index:
    - Commodity proportion: the corresponding merchants' types and the number of commodities released through the channel / the total number of commodities released through the corresponding channel
    - Dynamic distribution rate: commodity quantity of dynamic distribution / commodity quantity released
    - Rejection rate of first release: commodity quantity of rejection for the first release / commodity quantity released
    - Success rate of first release: commodity quantity of success for the first release / commodity quantity released
    - Proportion of commodities with success for the second time after the rejection for the first release: commodity quantity of success for the second time after the rejection for the first release / commodity quantity for the rejection of the first release
    - Proportion of commodities with success in final after the rejection for the first release: commodity quantity of success in final after the rejection for the first release / commodity quantity for the rejection of the first release
    - Proportion of commodities with failures in final after the rejection for the first release: commodity quantity of failure in final after the rejection for the first release / commodity quantity for the rejection of the first release
    - The waiting duration and the number of submissions for the final success after the rejection for the first release: in the commodities of final success after the rejection for

the first release:

- ✧ Average waiting duration: AVG (the creation time of the final examination corresponds to checked_time - the creation time of the first examination corresponds to sbmt_time)
- ✧ Average number of submissions: AVG (number of examinations)

2.Analytical process

Old and new merchants: the proportion of commodities released by new merchants through the merchants' APP is higher, there is little difference for the rejection of old and new merchants.
Dynamic distribution: the open platform has the most commodities released but the lowest dynamic distribution rate, the dynamic distribution rate of merchants' APP is the highest among all the channels.

Rejection: the overall rejection rate of first release through the channels is 10.75% which is relatively high (of which the rejection rate of the open platform is the highest), and the proportion of commodities with failures in final after the rejection of the first release reaches 97.05%.
The waiting duration and the number of submissions: the overall waiting duration of all the channels is 2.74 days (with little difference for the respective channels), the number of submissions is 4.53 which is relatively high (it is 5-6 times for merchants' APP and PC).

The current status for the rejection of commodities on the platform

Merchant type	Channel	Commodity quantity released	Commodity proportion	Dynamic distribution rate	Rejection rate of first release	Success rate of first release	After rejection of first release			Final success after rejection of first release	
							Proportion of commodities with success in second time	Proportion of commodities with success in final	Proportion of commodities with failure in final	Average waiting duration (days)	Average submission numbers
TOTAL	TOTAL	483,805,498	100.00%	29.31%	10.75%	74.98%	1.15%	1.65%	97.05%	2.74	4.53
	Open Platform	336,956,965	100.00%	22.45%	12.75%	76.30%	0.63%	0.95%	98.25%	2.79	3.32
	Merchants' APP	117,413,448	100.00%	47.70%	6.29%	73.35%	1.32%	1.71%	96.29%	2.75	6.10

	PC	29,435,085	100.00%	34.47%	5.63%	66.41%	13.80%	19.53%	69.20%	2.66	5.47
Old Merchants	TOTAL	249,824,878	51.64%	25.50%	10.98%	75.77%	1.41%	2.05%	96.67%	2.69	4.38
	Open Platform	187,908,850	55.77%	19.76%	12.50%	77.42%	0.73%	1.10%	98.15%	2.75	3.31
	Merchants' APP	45,387,376	38.66%	45.59%	6.78%	73.01%	1.74%	2.38%	95.43%	2.83	6.21
	PC	16,528,652	56.15%	35.60%	5.17%	64.64%	18.75%	26.82%	60.65%	2.57	5.05
New Merchants	TOTAL	233,980,620	48.36%	33.38%	10.50%	74.14%	0.86%	1.20%	97.46%	2.83	4.83
	Open Platform	149,048,115	44.23%	25.84%	13.06%	74.89%	0.51%	0.75%	98.38%	2.86	3.35
	Merchants' APP	72,026,072	61.34%	49.03%	5.98%	73.56%	1.01%	1.23%	96.90%	2.65	5.94
	PC	12,906,433	43.85%	33.01%	6.22%	68.69%	8.54%	11.75%	78.30%	2.89	6.48

IV. Field and requirement of attached data:
Idea target: checking the profit brought by this action

- Requirement and fact verification
  - Date of statistics: November 1 2021 to December 1 2021
  - Merchant coverage: non Daipai merchants
  - Channel coverage: PC, open platform and merchants' APP
  - Commodity coverage: newly submitted and released commodities (the creation time of the first examination corresponds to is_shop=0)

- Index requirement:
  - Rejected commodities of first release: the first examination time corresponds to check_status=3
  - Re-submitted commodities after the rejection of the first release: in the rejected

commodities of the first release, the number of submission >=2

- Commodities without re-submission after the rejection of the first release: in the rejected commodities of the first release, the number of submission =1
- Commodities without success in final after the rejection of the first release: the first examination time is November 1st, and the creation of the first examination corresponds to check_status=3, the final examination time corresponds to check_status IN (1,3)
- Failure in final after the rejection of the first release and there is no same commodity in the child category of the shop: in the commodities without success in final after the rejection of the first release, there is no same child category that can be purchased in the same shop on October 31
- Shops without re-submission after the rejection of the first release: the first examination time is November 1st, and the first examination time corresponds to check_status=3, the number of submission =1
- In the shops without re-submission after the rejection of the first release, shops release the same commodities in the child category within 3 days: shops without re-submission after the rejection of the first release re-submitted the commodities in the child category in the period from November 1st to November 3rd after the first examination of commodities.
- In the shops without re-submission after the rejection of the first release, shops release the commodities with the same title within 3 days: shops without re-submission after the rejection of the first release re-submitted the commodities with the same title in the period from November 1st to November 3rd after the first examination of commodities.

Stage Summary:

The proportion of commodities submitted again after the rejection of the first release is not high. The commodities with failure in final after the rejection are helpful for the abundance of commodities in shop. The merchants are accustomed to create commodities of rejection through the release of new commodities (the same category or the same title) for which we suggest to optimize the rejection link

Question Proposing:

1.1 Will the merchants submit again after the rejection of the first release?

The overall proportion of commodities with re-submission is 9.18% of which it is 36.98% for the PC channel.

Re-submission after the rejection of the first release

channel	Commodity quantity after the rejection of the first release	Commodity quantity with re-submission after the rejection of the first release	Proportion of commodities with re-submission
TOTAL	61394226	5638567	9.18%
Open platform	51381513	4031328	7.85%
Merchants' APP	8296132	926056	11.16%
PC	1716581	681183	39.68%

1.2 Will the rejected commodities be helpful for the improvement of abundance of commodities in shop?

The overall proportion of commodities that are not under the same child category in the same shop is 32.51%, the commodities without success in final after the rejection of the first release is helpful ti improve the abundance of commodities in shop.

1.3 Will the shops of rejection create the commodities of rejection through the release of new commodities in future?

As a whole, 75.08% of the shops of rejection will release the commodities in the same child category within 3 days, 73.72% of the shops of rejection will release the commodities with the same title within 3 days, the merchants have the habits of creating commodities of rejection through the release of new commodities.

Commodities of rejection created through the release of new commodities:

Channels	Number of shops without re-submission after the rejection of first release	Proportion of shops with the release of commodities in the same child category within 3 days	Proportion of shops with the release of commodities with the same title within 3 days
TOTAL	36913	75.08%	73.72%
Open Platform	20061	78.79%	77.06%
Merchants' APP	10677	72.18%	71.42%
PC	6175	68.02%	66.82%

I. Field and requirement of attached data

Sampling of commodities without re-submission after the rejection of the first release		Sampling of commodities created after rejection through the release of new commodities	
Field name	Field requirement	Field name	Field requirement
channel		channel	
Shop ID		Shop ID	Random sampling of 250 shops in each channel
Commodity ID	Random sampling of 250 commodities in each channel	Commodity ID without re-submission after the rejection of first release	Commodities submitted 10 times before the sampling of each shop
Commodity name of first release		Name of rejected commodity	
		Examination time	
		Commodity ID newly submitted after the rejection with the shop	

		Name of submitted commodity	
		Time of submission	

Demonstratration of the generation process of this table:

Current status of rejection for commodities on platform

Merchant type	Channel	Commodity quantity released	Commodity proportion	Dynamic distribution rate	Rejection rate of first release	Success rate of first release	After rejection of first release			Final success after rejection of first release	
							Proportion of commodities with success in second time	Proportion of commodities with success in final	Proportion of commodities with failure in final	Average waiting duration (days)	Average submission numbers
TOTAL	TOTAL	483,805,498	100.00%	29.31%	10.75%	74.98%	1.15%	1.65%	97.05%	2.74	4.53
	Open Platform	336,956,965	100.00%	22.45%	12.75%	76.30%	0.63%	0.95%	98.25%	2.79	3.32
	Merchants' APP	117,413,448	100.00%	47.70%	6.29%	73.35%	1.32%	1.71%	96.29%	2.75	6.10
	PC	29,435,085	100.00%	34.47%	5.63%	66.41%	13.80%	19.53%	69.20%	2.66	5.47
Old Merchants	TOTAL	249,824,878	51.64%	25.50%	10.98%	75.77%	1.41%	2.05%	96.67%	2.69	4.38
	Open Platform	187,908,850	55.77%	19.76%	12.50%	77.42%	0.73%	1.10%	98.15%	2.75	3.31
	Merchants	45,387,376	38.66%	45.59%	6.78%	73.01	1.74%	2.38%	95.43%	2.83	6.21

	’ APP					%					
	PC	16,528,652	56.15%	35.60%	5.17%	64.64%	18.75%	26.82%	60.65%	2.57	5.05
New Merchants	TOTAL	233,980,620	48.36%	33.38%	10.50%	74.14%	0.86%	1.20%	97.46%	2.83	4.83
	Open Platform	149,048,115	44.23%	25.84%	13.06%	74.89%	0.51%	0.75%	98.38%	2.86	3.35
	Merchants’ APP	72,026,072	61.34%	49.03%	5.98%	73.56%	1.01%	1.23%	96.90%	2.65	5.94
	PC	12,906,433	43.85%	33.01%	6.22%	68.69%	8.54%	11.75%	78.30%	2.89	6.48

In the process of data analysis, the python components I have utilized include: numpy package and pandas’ package

1. Commodity examination table (table a), including field commodity id, shop id, commodity examination time, commodity submission time, commodity examination time, commodity examination status, commodity submission channel

2. Shop registration table (table b), including the field, shop registration, time for the pass of examination, shop id

Process:

1. Define the channel for the submission of commodities: which channel it is for the first submission of commodity?
2. Need to divide the commodities in groups in accordance with the commodity id first, and sequence in accordance with the time of submission, take the submission channel corresponding to the commodity id of the earliest submission time, it requires the following functions:

   df.groupby( ).apply(**lambda** x: x.sort_values(, ascending=False))

3. Commodity quantity released: it requires the summary of segmentation for the commodity quantity: It requires the recision function drop_duplicates, and summarize through the function of df.groupby( ).count（）

Access demand for the rejection of commodities on platform(Conclusion):
Subscribed, Chu Gege, last modified by: Chu Gege November 26th 2022 at 21:38
Background:
It is expected to put the examination rules for some commodities at the postposition to preposition along with management to improve the success rate of commodities released by merchants and merchants' experience, the profit of this project will be initially evaluated at the early stage;
Demand:
Under the non-Daipai merchants, the following data:
1. Take the commodity quantity without success for the release in final after the rejection of commodities newly submitted and released in November (check_status=1, 3, the last pt) and the success rate of first release. There is no limitation for the specific time of rejection and examination of commodities, those that have not passed the examination will be considered as those fail to release for the treatment.
The total quantity of commodities with success for the new release in November, commodity quantity with orders to date and the total quantity of commodities for rejection (singularly check the commodity quantity with success for the release after the rejection for one time) and the total quantity of commodities without success for the release after the rejection, differentiate the bapp/mms/open channels for respective statistics.
Conduct extra statistics for the above stated data of new merchants, the requirement for new merchants is to register in October.
2. Take the average waiting duration of rejected commodities released in November and commodities with success for release in final
In the commodities with success for release in final and the rejected commodities released in November, calculate the average waiting duration of commodities, differentiate the bapp/mms/open channels for respective statistics.
3. Take the average submission times of commodities with success for release in final and the rejected commodities released in November:
In the commodities with success for release in final and the rejected commodities submitted and released in November, calculate the number of examinations in the average examination table, differentiate the bapp/mms/open channels for respective statistics.

Requirement:
Daipai shop: table goods.goods_oprty_gojp_mail_s_d, limit is_blk=0 and is_prom=0
Commodities newly submitted and released in November:
In table dwd_prod_goods_cmmt_i_d, limit pt>=20211101 and pt<20211201, and is_shop=0, and re-take goodsid;
Commodity quantity with orders:
Commodities in table dws_trde_mail_std_d are commodities with orders to date, take goodsid;
Rejection of commodities:
In table dwd_prod_goods_cmmt_i_d, check_status=3
Requirements for new merchants:
In table dwb_use_mall_bsc_s_d, fst_aprvd_time>="20211001"
Waiting duration of commodities:

The latest checked_time of the same commodity - the earliest sbmt_time

Supplemental feedback:
one data was also taken for the management from which it can be seen that many rejected commodities has a second submission rate of 0.03% only, merchants at the head of rejected commodities are also sampled from which it can be discovered that the shops of many punished merchants become invalid after the rejection. It indicates that many merchants are merchants of bad conduct who are releasing some bad commodities. Normal merchants will submit after the rejection of commodities, what needs to be confirmed is the proportion of repeated submission in reality which can be utilized to calculate the proportion of merchants of bad conduct. In addition, it may also be utilized to confirm if there are truly many merchants who release the rejected commodities again through the release of new commodities and determine if there is still a certain space for interaction design of rejection link.
1. Verify the second submission rate of 0.3% concluded by the management; take the proportion of commodities with second submission in commodities of first rejection in November (examination record >=2 pieces), the titles and mallid of commodities without second submission and commodities that are randomly rejected, differentiate the channels:
Requirement: earliest pt check_status=3, the record of goodsid to date >=2 pieces
Field: quantity of commodities that are first rejected in November, of which there is commodity quantity with second submission
2. Check the improvement of the commodity abundance in shops by the rejected commodities in reality, verify if the rejected commodity basically improves the commodity abundance of shops: Take the commodity quantity without success for the release in final after the first rejection on November 1st, check the commodities in shops before October 31st to determine if there are commodities under the same child category, take the quantity of commodities in shops that are not under the same child category of rejected commodities on October 31st, differentiate the channels:
Requirement: earliest pt check_status=3, latest pt check_status=1, 3, cate4_id of commodities under match table dwb_prod_goods_s_d to match the commodities with difference so as to do the statistics for the commodity quantity;
Fields included: commodity quantity without success for the release after rejection, commodity quantity that are not under the same child category within the shop
3. Check if the shops corresponding to the commodities without second submission after rejection create rejected commodities through the release of new commodities in future, verify the need for the optimization of the rejection link: Check the commodity quantity without second submission after the first rejection on November 1st, take the proportion of commodities under the same child category released within 3 days by the shop after the rejection, take the proportion of commodities with the same title released within 2 days by the shop after the rejection, differentiate the channels;

Appendix:

```
--Author: Jiale Chen
--Name:Analysis of reasons for low secondary submission rate after rejection of the first release of goods
--DateTime:20220126
--Description:
--ModifyBy:
--ModifyDate:
--ModifyDesc:
--Copyright pinduoduo.com


--------------------------create table--------------------------
#*
CREATE TABLE .
(

)
COMMENT ''
PARTITIONED BY (pt STRING COMMENT )
STORED AS ORC
;
*#


-----------------------SQL processing-----------------------
-- 1.Parameter setting
-- SET tez.grouping.min-size=1024000000;
-- SET tez.grouping.max-size=2048000000;
-- SET mapred.max.split.size=2048000000;
-- SET mapred.min.split.size=2048000000;
-- SET mapreduce.job.reduces=3000;
-- SET hive.tez.container.size=4000;
SET hive.exec.orc.split.strategy=BI;
SET tez.grouping.split-waves=0;
SET hive.map.aggr = true;
SET hive.groupby.skewindata=true;

-- 2.table processing
DROP TABLE IF EXISTS jialechen.gauss_85942_${env.YYYYMMDD8}_1;
CREATE TABLE jialechen.gauss_85942_${env.YYYYMMDD8}_1 AS
SELECT
 t1.*
FROM ( SELECT
       pt
      ,ROW_NUMBER() OVER(PARTITION BY goods_id ORDER BY create_time ASC) AS rn
      ,mall_id
      ,goods_id
      ,goods_name
      ,cate_id
      ,CASE WHEN src = 0 THEN 'PC'
            WHEN src = 1 THEN '开放平台'--open site
            WHEN src = 2 THEN '商家APP'--platform APP
            ELSE '其他'
            END AS src
      ,check_status
      ,checked_time
      ,sbmt_time
      ,create_time
      ,is_shop
      FROM dwd.dwd_prod_goods_cmmt_i_d
      WHERE pt BETWEEN '2021-11-01' AND '2021-12-01'
    ) t1
LEFT JOIN ( SELECT
             mall_id
           FROM goods.goods_oprty_gojp_mall_s_d
           WHERE pt = '${env.YYYYMMDD}'
           AND is_blk = 0
           AND is_prom = 0
           GROUP BY mall_id
         ) t2
ON t1.mall_id = t2.mall_id
WHERE t2.mall_id IS NULL
;

-- 3.summary
-- 1
SELECT
 NVL(fst_src,'TOTAL') `渠道`--channel
,SUM(1) `首次发布驳回商品数`--Number of rejected products for first release
,SUM(IF(cnt>=2,1,0)) `首次发布驳回并再次提交商品数`--First release rejected and resubmitted item count
FROM ( SELECT
       mall_id
      ,goods_id
      ,MIN(ARRAY(rn,src))[1] AS fst_src -- 首次的渠道
      ,MIN(ARRAY(rn,is_shop))[1] AS fst_is_shop -- 首次是否是线上商品
      ,MIN(ARRAY(rn,check_status))[1] AS fst_check_status -- 首次的审核状态
      ,MAX(rn) AS cnt
      FROM jialechen.gauss_85942_${env.YYYYMMDD8}_1
      GROUP BY
       mall_id
      ,goods_id
    ) t
WHERE fst_src <> '其他'--others
AND fst_is_shop = 0
AND fst_check_status = 3
GROUP BY fst_src
```

```
WITH CUBE
ORDER BY `渠道`--channels
LIMIT 100
;
SELECT
 fst_src `渠道`--channel
,mall_id `店铺ID`--store ID
,goods_id `商品ID`--good ID
,fst_goods_name `首次发布的商品名称`--name for the publisher
FROM ( SELECT
       fst_src
      ,mall_id
      ,goods_id
      ,fst_goods_name
      ,ROW_NUMBER() OVER(PARTITION BY fst_src ORDER BY RAND()) AS rn
      FROM ( SELECT
             mall_id
            ,goods_id
            ,MIN(ARRAY(rn,goods_name))[1] AS fst_goods_name
            ,MIN(ARRAY(rn,src))[1] AS fst_src
            ,MIN(ARRAY(rn,is_shop))[1] AS fst_is_shop
            ,MIN(ARRAY(rn,check_status))[1] AS fst_check_status
            ,MAX(rn) AS cnt
            FROM jialechen.gauss_85942_${env.YYYYMMDD8}_1
            GROUP BY
             mall_id
            ,goods_id
           ) rn
      WHERE fst_src <> '其他'--others
      AND fst_is_shop = 0
      AND fst_check_status = 3
      AND cnt = 1
     ) t
WHERE rn <= 250
ORDER BY `渠道`,`商品ID`--channel, good ID
LIMIT 100000
;


-- 2
SELECT
 NVL(t1.fst_src,'TOTAL') `渠道`--channel
,SUM(1) `首次发布驳回后最终未成功商品数`--Number of products that failed after initial launch rejection
,SUM(IF(t2.cate_id IS NULL,1,0)) `首次发布驳回后最终未成功,且店铺内无该叶子类目商品数`--The first release was rejected and failed, and there was no product number
FROM ( SELECT
       *
      FROM ( SELECT
             mall_id
            ,goods_id
            ,MIN(ARRAY(rn,pt))[1] AS fst_pt
            ,MIN(ARRAY(rn,cate_id))[1] AS fst_cate_id
            ,MIN(ARRAY(rn,src))[1] AS fst_src
            ,MIN(ARRAY(rn,is_shop))[1] AS fst_is_shop
            ,MIN(ARRAY(rn,check_status))[1] AS fst_check_status
            ,MAX(ARRAY(rn,check_status))[1] AS lst_check_status
            ,MAX(rn) AS cnt
            FROM jialechen.gauss_85942_${env.YYYYMMDD8}_1
            GROUP BY
             mall_id
            ,goods_id
           ) rn
      WHERE fst_pt = '2021-11-01'
      AND fst_src <> '其他'--others
      AND fst_is_shop = 0 -- 非线上商品
      AND fst_check_status = 3 -- 首次发布驳回
      AND lst_check_status IN (1,3) -- 末次发布未成功
     ) t1
LEFT JOIN ( SELECT
            mall_id
           ,cate_id
           FROM dwb.dwb_prod_goods_info_s_d
           WHERE pt = '2021-10-31'
           AND is_avlb_buy = 1
           GROUP BY
            mall_id
           ,cate_id
          ) t2
ON t1.mall_id = t2.mall_id
AND t1.fst_cate_id = t2.cate_id
GROUP BY t1.fst_src
WITH CUBE
ORDER BY `渠道`--channel
LIMIT 100
;


-- 3
SELECT
 NVL(t1.fst_src,'TOTAL') `渠道`--Total channel
,COUNT(DISTINCT t1.mall_id) `首次发布驳回且未再次提交店铺数`--The first release was rejected and the number of stores was not submitted again
,COUNT(DISTINCT IF(t2.cate_id IS NOT NULL AND t2.sbmt_time>t1.fst_checked_time,t1.mall_id,NULL)) `首次发布驳回且未再次提交店铺中，3天内发布同叶子类目商品的店铺数`
--Number of stores that rejected the first release and did not submit the product again within 3 days
,COUNT(DISTINCT IF(t3.goods_name IS NOT NULL AND t2.sbmt_time>t1.fst_checked_time,t1.mall_id,NULL)) `首次发布驳回且未再次提交店铺中，3天内发布同标题商品的店铺数`
--Number of stores that rejected the first release and did not submit the product again within 3 days
FROM ( SELECT
```

```
    *
    FROM ( SELECT
        *
        ,ROW_NUMBER() OVER(PARTITION BY mall_id ORDER BY fst_checked_time ASC) AS rn
        FROM ( SELECT
            mall_id
            ,goods_id
            ,MIN(ARRAY(rn,pt))[1] AS fst_pt
            ,MIN(ARRAY(rn,checked_time))[1] AS fst_checked_time
            ,MIN(ARRAY(rn,cate_id))[1] AS fst_cate_id
            ,MIN(ARRAY(rn,goods_name))[1] AS fst_goods_name
            ,MIN(ARRAY(rn,src))[1] AS fst_src
            ,MIN(ARRAY(rn,is_shop))[1] AS fst_is_shop
            ,MIN(ARRAY(rn,check_status))[1] AS fst_check_status
            ,MAX(rn) AS cnt
            FROM jialechen.gauss_85942_${env.YYYYMMDD8}_1
            GROUP BY
            mall_id
            ,goods_id
          ) n
        WHERE fst_pt = '2021-11-01'
        AND TO_DATE(from_unixtime(fst_checked_time,'yyyy-MM-dd HH:mm:ss')) = '2021-11-01'
        AND fst_src <> '其他'
        AND fst_is_shop = 0
        AND fst_check_status = 3
        AND cnt = 1
      ) m
    WHERE rn = 1
  ) t1
LEFT JOIN ( SELECT
        mall_id
        ,cate_id
        ,sbmt_time
        FROM jialechen.gauss  85942  ${env.YYYYMMDD8}  1
        WHERE is_shop = 0
        AND TO_DATE(from_unixtime(sbmt_time,'yyyy-MM-dd HH:mm:ss')) BETWEEN '2021-11-01' AND '2021-11-03'
      ) t2
ON t1.mall_id = t2.mall_id
AND t1.fst_cate_id = t2.cate_id
LEFT JOIN ( SELECT
        mall_id
        ,goods_name
        ,sbmt_time
        FROM jialechen.gauss_85942_${env.YYYYMMDD8}_1
        WHERE is_shop = 0
        AND TO_DATE(from_unixtime(sbmt_time,'yyyy-MM-dd HH:mm:ss')) BETWEEN '2021-11-01' AND '2021-11-03'
      ) t3
ON t1.mall_id = t3.mall_id
AND t1.fst_goods_name = t3.goods_name
GROUP BY t1.fst_src
WITH CUBE
ORDER BY `渠道`--channel
LIMIT 100
;

-- 4
SELECT
 src `渠道`--channel
,mall_id `店铺ID`--store name
,rjt_goods_id `首次发布驳回且未再次提交的商品ID`--item ID rejected for the first time and not resubmitted
,rjt_goods_name `驳回商品的名称`--name of the rejected good
,rjt_checked_time `审核时间`--time for inspection
,new_goods_id `同店驳回后新提交的商品ID`--New product ID submitted after same-store rejection
,new_goods_name `提交商品的名称`--Submit the name of the commodity
,new_sbmt_time `提交时间`--time for submission
FROM ( SELECT
    *
    ,ROW_NUMBER() OVER(PARTITION BY src,mall_id ORDER BY new_sbmt_time ASC) AS rn
    FROM ( SELECT
        t1.fst_src AS src
        ,t1.mall_id
        ,from_unixtime(t1.fst_checked_time,'yyyy-MM-dd HH:mm:ss') AS rjt_checked_time
        ,t1.goods_id AS rjt_goods_id
        ,t1.fst_goods_name AS rjt_goods_name
        ,from_unixtime(t2.sbmt_time,'yyyy-MM-dd HH:mm:ss') AS new_sbmt_time
        ,t2.goods_id AS new_goods_id
        ,t2.goods_name AS new_goods_name
        FROM ( SELECT
            *
            FROM ( SELECT
                *
                ,ROW_NUMBER() OVER(PARTITION BY fst_src ORDER BY RAND()) AS rn_2
                FROM ( SELECT
                    *
                    FROM ( SELECT
                        *
                        ,ROW_NUMBER() OVER(PARTITION BY mall_id ORDER BY fst_checked_time ASC) AS rn_1
                        FROM ( SELECT
                            mall_id
                            ,goods_id
                            ,MIN(ARRAY(rn,pt))[1] AS fst_pt
                            ,MIN(ARRAY(rn,checked_time))[1] AS fst_checked_time
                            ,MIN(ARRAY(rn,cate_id))[1] AS fst_cate_id
                            ,MIN(ARRAY(rn,goods_name))[1] AS fst_goods_name
                            ,MIN(ARRAY(rn,src))[1] AS fst_src
```

```
                                ,MIN(ARRAY(m,is_shop))[1] AS fst_is_shop
                                ,MIN(ARRAY(m,check_status))[1] AS fst_check_status
                                ,MAX(m) AS cnt
                                FROM jialechen.gauss_85942_${env.YYYYMMDD8}_1
                                GROUP BY
                                 mall_id
                                ,goods_id
                               ) nn
                              WHERE fst_pt = '2021-11-01'
                              AND TO_DATE(from_unixtime(fst_checked_time,'yyyy-MM-dd HH:mm:ss')) = '2021-11-01'
                              AND fst_src <> '其他'--others
                              AND fst_is_shop = 0 -- 非线上商品
                              AND fst_check_status = 3
                              AND cnt = 1
                             ) n
                            WHERE m_1 = 1
                           ) mm
                         ) m
                        WHERE m_2 <= 250
                       ) t1
                LEFT JOIN ( SELECT
                             mall_id
                            ,goods_id
                            ,goods_name
                            ,sbmt_time
                            FROM jialechen.gauss_85942_${env.YYYYMMDD8}_1
                            WHERE is_shop = 0
                           ) t2
                ON t1.mall_id = t2.mall_id
                WHERE t2.mall_id IS NULL
                OR (t2.mall_id IS NOT NULL AND t2.sbmt_time > t1.fst_checked_time)
                GROUP BY
                 t1.fst_src
                ,t1.mall_id
                ,from_unixtime(t1.fst_checked_time,'yyyy-MM-dd HH:mm:ss')
                ,t1.goods_id
                ,t1.fst_goods_name
                ,from_unixtime(t2.sbmt_time,'yyyy-MM-dd HH:mm:ss')
                ,t2.goods_id
                ,t2.goods_name
               ) tt
    ) t
WHERE m <= 10
ORDER BY `渠道`,`店铺ID`,`提交时间`--'channel', 'store ID', 'submission time'
LIMIT 100000
;

-- 4.recycle
DROP TABLE IF EXISTS jialechen.gauss_85942_${env.YYYYMMDD8}_1;
```

```
--Author: Jiale Chen
--Name:Rejection of platform commodities
--DateTime:20220120
--Description:
--ModifyBy:
--ModifyDate:
--ModifyDesc:
--Copyright pinduoduo.com


--------------------------create table--------------------------
#*
CREATE TABLE .
(

)
COMMENT ''
PARTITIONED BY (pt STRING COMMENT '日期分区') #The date of the partition--time distinction
STORED AS ORC
;
'*#
----------------------------------------------------




--------------------------SQL processing--------------------------
-- 1.Parameter setting
-- SET tez.grouping.min-size=1024000000;
-- SET tez.grouping.max-size=2048000000;
-- SET mapred.max.split.size=2048000000;
-- SET mapred.min.split.size=2048000000;
-- SET mapreduce.job.reduces=3000;
-- SET hive.tez.container.size=4000;
SET hive.exec.orc.split.strategy=BI;
SET tez.grouping.split-waves=0;
SET hive.map.aggr = true;
SET hive.groupby.skewindata=true;

-- 2.Temporary table processing


-- 3.summary
SELECT
 NVL(IF(t3.mall_id IS NOT NULL,'新商家','老商家'),'TOTAL') `商家类型`--New merchant ',' old merchant '),'TOTAL') 'merchant type'
,NVL(t1.fst_src,'TOTAL') `渠道`--TOTAL channel
,COUNT(t1.goods_id) `发布商品数`--Number of publisher goods
,COUNT(t4.goods_id) `动销商品数`--Number of moving pin goods
,COUNT(t4.goods_id)/COUNT(t1.goods_id) `动销率`--rate of moving pin
,SUM(IF(t1.fst_check_status=3,1,0)) `首次发布驳回商品数`--Number of rejected products for first release
,SUM(IF(t1.fst_check_status=3,1,0))/COUNT(t1.goods_id) `首次发布驳回率`--Rejection rate of first release
,SUM(IF(t1.fst_check_status=2,1,0)) `首次发布成功商品数`--Number of successful products released for the first time
,SUM(IF(t1.fst_check_status=2,1,0))/COUNT(t1.goods_id) `首次发布成功率`--Success rate of first release

,SUM(IF(t1.fst_check_status=3 AND t1.scnd_check_status=2,1,0)) `首次发布驳回后二次成功商品数`-- Number of second successful products after first release rejection
,SUM(IF(t1.fst_check_status=3 AND t1.scnd_check_status=2,1,0))/SUM(IF(t1.fst_check_status=3,1,0)) `首次发布驳回后二次成功商品占比`--Proportion of second successful products after first release rejection
,SUM(IF(t1.fst_check_status=3 AND t1.lst_check_status=2,1,0)) `首次发布驳回后最终成功商品数`--Final number of successful products after initial launch rejection
,SUM(IF(t1.fst_check_status=3 AND t1.lst_check_status=2,1,0))/SUM(IF(t1.fst_check_status=3,1,0)) `首次发布驳回后最终成功商品占比`--Percentage of final successful products after initial launch rejection
,SUM(IF(t1.fst_check_status=3 AND t1.lst_check_status IN (1,3),1,0)) `首次发布驳回后最终未成功商品数`--Number of products that failed after initial launch rejection
,SUM(IF(t1.fst_check_status=3 AND t1.lst_check_status IN (1,3),1,0))/SUM(IF(t1.fst_check_status=3,1,0)) `首次发布驳回后最终未成功商品占比`--Percentage of unsuccessful products after initial rejection

,AVG(IF(t1.fst_check_status=3 AND t1.lst_check_status=2,t1.lst_checked_time-t1.fst_sbmt_time,NULL)) `首次发布驳回后最终成功的平均等待时长`--Average waiting time after initial release rejection for final success
,AVG(IF(t1.fst_check_status=3 AND t1.lst_check_status=2,t1.cnt,NULL)) `首次发布驳回后最终成功的平均提交次数`--The average number of successful submissions after a first release rejection
FROM ( SELECT
       mall_id
      ,goods_id
      ,MIN(ARRAY(m,src))[1] AS fst_src
      ,MIN(ARRAY(m,is_shop))[1] AS fst_is_shop
      ,MIN(ARRAY(m,check_status))[1] AS fst_check_status
      ,MAX(IF(rn=2,check_status,NULL)) AS scnd_check_status
      ,MAX(ARRAY(m,check_status))[1] AS lst_check_status
      ,MIN(ARRAY(m,sbmt_time))[1] AS fst_sbmt_time
      ,MAX(ARRAY(m,checked_time))[1] AS lst_checked_time
      ,MAX(rn) AS cnt
      FROM ( SELECT
            ROW_NUMBER() OVER(PARTITION BY goods_id ORDER BY create_time ASC) AS m
            ,mall_id
            ,goods_id
            ,CASE WHEN src = 0 THEN 'PC'
                WHEN src = 1 THEN '开放平台'--open site
                WHEN src = 2 THEN '商家APP'--B side APP
                ELSE '其他'--others
                END AS src
            ,check_status
            ,checked_time
            ,sbmt_time
            ,is_shop
            FROM dwd.dwd_prod_goods_cmmt_i_d
            WHERE pt BETWEEN '2021-11-01' AND '2021-12-01'
          ) m
```

```
      GROUP BY
       mall_id
      ,goods_id
    ) t1
LEFT JOIN ( SELECT
             mall_id
           FROM goods.goods_oprty_gojp_mall_s_d
           WHERE pt = '${env.YYYYMMDD}'
           AND is_blk = 0
           AND is_prom = 0
           GROUP BY mall_id
         ) t2
ON t1.mall_id = t2.mall_id
LEFT JOIN ( SELECT
             mall_id
           FROM dwb.dwb_usr_mall_bsc_s_d
           WHERE pt = '${env.YYYYMMDD}'
           AND TO_DATE(fst_aprvd_time) >= '2021-10-01'
         ) t3
ON t1.mall_id = t3.mall_id
LEFT JOIN ( SELECT
              goods_id
           FROM dws.dws_trde_goods_std_d
           WHERE pt = '${env.YYYYMMDD}'
           AND ordr_crt_ordr_cnt_std > 0 -- 历史截至当日下单订单数>0
         ) t4
ON t1.goods_id = t4.goods_id
WHERE t2.mall_id IS NULL
AND t1.fst_src <> '其他'--others
AND t1.fst_is_shop = 0 -- 非线上商品
GROUP BY
 IF(t3.mall_id IS NOT NULL,'新商家','老商家')--new customers/old customers
,t1.fst_src
WITH CUBE
ORDER BY `商家类型`,`发布商品数` DESC--Merchant type ', 'number of publisher items
LIMIT 100
;

-- recycle
```