# Data Description

The data consists of 5 tables, and ranges from 2017/01/01 to 2017/07/31. We included more than 200 merchants using Cainiao's integrated warehousing and fulfillment service, and more than 200 merchants who managed their own inventory and fulfillment.

The first table is the detailed information about customer orders on Tmall for those merchants. The description of the schema is as follows:

1. customer_order_data

| Field | Data Type | Description | Sample |
|---|---|---|---|
| day | string | Date of payment | 20170101 |
| order_id | string | ID of the order | 123456 |
| item_det_info | string | Includes item id, quantity and payment amount for all items in the order | 123:2:4.5;234:2:35.5. Note that different items are separated by ";", and the information for each item is separated by ":". The meaning of the fields are item id, quantity, payment amount. Note that the payment amount is the total payment for all quantity of this item. |
| pay_timestamp | string | Time of payment (rounded to 1 mins intervals) | 2017-01-01 08:50:00<br>All timestamp are based on Beijing time. |
| buyer_id | string | ID of the user who placed the order | 111 |
| promise_speed | int | Promised delivery speed | 1: same day delivery, 2: next day delivery, 0 or null :no promise |
| if_cainiao | int | Whether the order is delivered from cainiao's warehouse, i.e., did this order use Cainiao's warehousing and delivery service | 0: not delivered from cainiao warehouse, 1: delivered from cainiao warehouse |

| | | | |
|---|---|---|---|
| merchant_id | string | ID of the merchant | 222 |
| Logistics_review_score | int | User's review score on the logistic service | Scale from 1 to 5, and higher score represents better logistic service |

The second table is the detailed inventory history. Note that we only have data for merchants using Cainiao's warehousing and fulfillment service. In this table, besides total inventory position, we also included replenishment and transfer inventory information. Note that besides replenishment and transfer, there are still other events which would alter the inventory position, such as return, lost inventory, etc. To simplify the matter, we do not include those detailed events since they are relatively rare. These rare events will result in the data has few percentage discrepancy. The description of the schema is as follows:

2. inventory_data

| Field | Data type | Description | Sample |
|---|---|---|---|
| day | string | Date of the inventory snapshot | 20170101 |
| item_id | string | ID of item | 123 |
| warehouse_id | int | ID of the warehouse | 1 |
| warehouse_city_id | int | City id of the warehouse | 22 |
| total_begin_qty | int | Total item quantity at the beginning of the day | 300 |
| total_end_qty | int | Total item quantity at the end of the day | 400 |
| Replen_in_qty | int | Total replenishment inventory received | 100 |
| transfer_in_qty | int | Total transfer-in inventory received | 100 |
| sale_out_qty | int | Total sale-out inventory | 100 |
| transfer_out_qty | int | Total transfer-out inventory sent | 100 |

The third table is about the meta-data for each item, including merchant, brand, and category information. The description of the schema is as follows:

3. item_view_data

| Field | Data type | Description | Sample |
|---|---|---|---|
| date | string | Date of item view | 20170101 |
| item_id | string | ID of the item | 123 |
| front_page_item_id | string | ID of the item at the e-commerce website | 456 |
| merchant_id | int | ID of the merchant | 12 |
| brand_id | int | ID of the brand of the item | 11 |
| category_id | int | ID of the main category of the item | 1 |
| sub_category_id | int | ID of the sub category of the item | 2 |
| pc_pv | int | Total page view on pc | 100 |
| app_pv | int | Total unique visitor count on app | 70 |
| pc_uv | int | Total unique visitor count on pc | 50 |
| app_uv | int | Total page view on app | 150 |
| if_cainiao | boolean | Whether the item is delivered by Cainiao's warehousing and fulfillment service | 1: Use Cainiao's warehouse, 0: not use Cainiao's warehouse |

The fourth table is about the aggregated online traffic and review information for each merchant, aggregated at the sub category level. The description of the schema is as follows:

4. merchant_category_data

| Field | Data type | description | sample |
|---|---|---|---|
| day | string | Date of the snapshot | 20170101 |
| merchant_id | string | ID of the merchant | 5 |
| subcategory_id | int | ID of the sub category | 1 |
| pc_pv | int | Total page view on pc | 100 |
| pc_uv | int | Total unique visitor count on pc | 20 |
| app_pv | int | Total page view on app | 150 |
| app_uv | int | Total unique visitor count on app | 30 |
| avg_logistic_review_score | double | Average review score on logistic service | 4.5（range in 1-5） |
| avg_order_quality_score | double | Average review score on order quality. Here, order quality means the quality of products in the order | 4.4（range in 1-5） |
| avg_service_quality_score | double | Average review score on service quality. Here service quality means the quality of online purchase service | 4.6（range in 1-5） |
| if_cainiao | boolean | Whether the merchant uses Cainiao’s warehousing and fulfillment service | 1: Use Cainiao’s warehouse, 0: not use Cainiao’s warehouse |

The fifth table is about the detailed information on the fulfillment logistics, including the carrier selected, the dispatch of the package, the arrival and departure of the package at each facility of the logistics network, and the acceptance of the package by the customer. Please note that this table also includes the orders that do not use Cainiao's warehouse and fulfillment service. The description of the schema is as follows:

5. logistics_detail_data

| Field | Data type | Description | Sample |
|---|---|---|---|
| order_id | string | ID of the purchase order | 12345 |
| order_date | string | Date of the payment of the order | 20170101 |
| logistic_order_id | string | ID of the logistic order | 123451 |
| action | string | CONSIGN: dispatch of the package from warehouse/merchant<br>GOT: package received by the carrier<br>ARRIVAL: arrive at the facility<br>DEPARTURE: leave the facility<br>SENT_SCAN: package left the delivery station<br>SIGNED: the buyer received the package<br>TRADE_SUCCESS: confirm the order is delivered<br>FAILURE: fail to deliver the order | arrival |
| facility_id | int | The facility that handles the package at this particular event. | 12 |
| facility_type | int | Type of the facility.<br>1: delivery station<br>2. logistic transfer center<br>Null: no facility or facility type not categorized | 1 |
| city_id | int | ID of the city | 1 |
| logistic_company_id | int | ID of the company that delivers the package | 1 |
| timestamp | string | Time that the event happened | 2017-01-01 08:50:00 |

Data_1.csv to data_7.csv contains the order and logistic data of each month and data_8.csv contains the data of item, inventory and seller.