

# Dataset Annexure

Column Name	Data Type	Column Description
b2c_c2c	String	Defines whether the transaction belongs to a business seller; possible values are B2C (business to consumer) or C2C (consumer to consumer).
seller_id	Long	Anonymized seller ID, used to cluster transactions from the same seller.
declared_handling_days	Integer	Handling days is the time taken by the seller to package the item(s) before shipping and to deliver the shipment to the carrier. Carrier acceptance marks the end of the handling time period. Seller claims a handling time period, which might/might not be the same as the actual handling time. Weekend days and holidays are excluded from handling days.
acceptance_scan_timestamp	Timestamp	Timestamp by the carrier accepting or first processing the package, in the local timezone of the carrier's acceptance facility. The time zone, when available, appears at the end of the entry.
shipment_method_id	Integer	Anonymized shipping service ID which signifies the type of the service declared by the seller to be used for shipping the item(s).
shipping_fee	Double	Shipping and handling fee charged by the seller for shipping of the item(s), in USD.
carrier_min_estimate	Integer	Carrier provided minimum transit time in the number of carrier working days for the given service.
carrier_max_estimate	Integer	Carrier provided maximum transit time in the number of carrier working days for the given service.

item_zip	String	US postal code giving the source/origination of the package; might be same as the seller location; might have different values for different transactions for the same seller for multi-locations sellers.
buyer_zip	String	US postal code of the ship-to/destination of the package; buyer specified address.
category_id	Integer	Anonymized broader level category for classifying type of item purchased.
item_price	Double	Price per item in a transaction, in USD.
quantity	Integer	Quantity of the items bought in a transaction.
payment_datetime	Timestamp	Payment time of the transaction, in the local timezone of the item location. The time zone, when available, appears at the end of the entry.
delivery_date	Date	Actual delivery date (YYYY-MM-DD) of the package, in the local timezone of the buyer.
weight	Double	Weight of the package in weight units.
weight_units	Integer	Units for weights. 1 – “lbs”, 2 – “kg”
package_size	String	Symbolic size of the package, examples: LETTER, LARGE_ENVELOPE, etc.
record_number	Integer	Unique record number.

## Additional Information

- There are certain irregularities in the data, some of them affecting only the training set, some only the quiz/test sets, or both. How frequent any of these irregularities occurs may also differ between training and quiz/test sets. Quiz and test sets have the same characteristics, so either an irregularity applies to both and in approximately the same frequency, or applies to neither quiz set nor test set. The training set and the eval/test sets cover different time periods. Eval and test sets cover the same time period.

- **carrier\_min\_estimate** and **carrier\_max\_estimate** have records with negative values indicating missing values.
- **declared\_handling\_days** is missing for some records.
- **payment\_datetime**, **acceptance\_scan\_timestamp**, and **delivery\_date** are in local time, in order: item local time, seller local time (usually but not always the same as carrier local time), and buyer local time.
- **payment\_datetime** and **acceptance\_scan\_timestamp** may have the seconds (minutes, hours) set to zero if that information is not available.
- **payment\_datetime** is in YYYY-MM-DD HH:MM:SS.SSSS [-+]HH:MM format where the [-+]HH:MM is the GMT offset.
- **acceptance\_scan\_timestamp** is in YYYY-MM-DD HH:MM:SS.SSSS [-+]HH:MM format where the [-+]HH:MM is the GMT offset.
- **delivery\_date** is in YYYY-MM-DD format.
- Seller handling payment cutoff time is assumed to be 2PM local time. This means, for example, if a seller specifies 0 handling days, then if payment occurs on a business day before 2PM in the seller's timezone then the seller should ship the item(s) the same day, if payment occurs after 2PM then the seller can ship the item(s) the next business day.
- **seller\_id** and **service\_id** may have records in the quiz and test set which have values not occurring in the training set.
- **package\_size** has the value NONE to indicate missing data.
- **weight** uses zero to indicate missing values, there may be some invalid values also. Some flat rate shipping services may default to always setting the weight to zero. A flat rate shipping service is one where the shipping cost charged by the carrier does not depend on the item weight.
- Both **item\_zip** code and **buyer\_zip** codes are assumed to be US postal codes (including territories). Several formats may be present, including 5 digit, 5+4 digit, and 9 digit (same meaning as 5+4 digit but without the dash in between). Entries may not be valid.
- There may be records where the (payment / acceptance / delivery) dates are inconsistent, for example the delivery date might be before the payment date.