Insta cart data dictionary

Link: https://gist.github.com/jeremystan/c3b39d947d9b88b3ccff3147dbcf6c6b#file-data\_description-md

Citation: The Instacart Online Grocery Shopping Dataset 2017”, Accessed from www.instacart.com/datasets/grocery-shopping-2017 via Kaggle on

1. order\_id: order identifier

2. user\_id: customer identifier

3. eval\_set: which evaluation set this order belongs in (see SET described below)

4. order\_number: the order sequence number for this user (1 = first, n = nth)

5. order\_dow: the day of the week the order was placed on

6. order\_hour\_of\_day: the hour of the day the order was placed on

7. days\_since\_prior: days since the last order, capped at 30 (with NAs for order\_number = 1)

products (50k rows):

• product\_id: product identifier

• product\_name: name of the product

• aisle\_id: foreign key

• department\_id: foreign key

aisles (134 rows):

• aisle\_id: aisle identifier

• aisle: the name of the aisle

deptartments (21 rows):

• department\_id: department identifier

• department: the name of the department

order\_products\_\_SET (30m+ rows):

• order\_id: foreign key

• product\_id: foreign key

• add\_to\_cart\_order: order in which each product was added to cart

• reordered: 1 if this product has been ordered by this user in the past, 0 otherwise

where SET is one of the four following evaluation sets (eval\_set in orders):

• "prior": orders prior to those users most recent order (~3.2m orders)

• "train": training data supplied to participants (~131k orders)

• "test": test data reserved for machine learning competitions (~75k orders)