The Instacart Online Grocery Shopping Dataset 2017 Data Descriptions

[Raw](https://gist.github.com/jeremystan/c3b39d947d9b88b3ccff3147dbcf6c6b/raw/0a6a6b96c80016cd1e44d8963745ce66196245e0/data_description.md)

[**data\_description.md**](https://gist.github.com/jeremystan/c3b39d947d9b88b3ccff3147dbcf6c6b#file-data_description-md)

orders (3.4m rows, 206k users):

* order\_id: order identifier
* user\_id: customer identifier
* eval\_set: which evaluation set this order belongs in (see SET described below)
* order\_number: the order sequence number for this user (1 = first, n = nth)
* order\_dow: the day of the week the order was placed on
* order\_hour\_of\_day: the hour of the day the order was placed on
* 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 that 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)

A screenshot of a computer

Description automatically generated