Description of Instacart Transaction Data

`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)