



Instacart Data Dictionary

This dataset was originally provided by Instacart as part of the 2017 Online Grocery Shopping Dataset. It contains over **34 million records** across multiple tables, including customer orders, products, departments, and aisles.



Orders (3.4M rows, 206K users)

Column Name	Description
`order_id`	Unique order identifier
`user_id`	Unique customer identifier
`eval_set`	Evaluation set: `prior`, `train`, or `test`
`order_number`	Sequence number of the order for each user
`order_dow`	Day of the week the order was placed
`order_hour_of_day`	Hour of the day the order was placed
`days_since_prior`	Days since the previous order (NA for first order)



Products (50K rows)

Column Name	Description
`product_id`	Unique product identifier
`product_name`	Name of the product
`aisle_id`	Foreign key to aisle table
`department_id`	Foreign key to department table



Aisles (134 rows)

Column Name	Description
`aisle_id`	Unique aisle identifier
`aisle`	Name of the aisle

Departments (21 rows)

Column Name	Description
`department_id`	Unique department identifier
`department`	Name of the department

Order_Products__SET (30M+ rows)

Column Name	Description
`order_id`	Foreign key to orders table
`product_id`	Foreign key to products table
`add_to_cart_order`	Sequence in which product was added to cart
`reordered`	1 if product was previously ordered by user, 0 otherwise

Evaluation Sets

The `eval_set` column in the orders table refers to one of the following:

- `prior`: All orders before the user's most recent order (~3.2M orders)
- `train`: Orders used for training (~131K orders)
- `test`: Orders reserved for prediction (~75K orders)

Citation

"The Instacart Online Grocery Shopping Dataset 2017". Originally accessed via Kaggle and distributed through my data analytics training program. This project is for educational purposes only and does not represent Instacart or its business operations."