Whether you shop from meticulously planned grocery lists or let whimsy guide your grazing, our unique food rituals define who we are. Instacart, a grocery ordering and delivery app, aims to make it easy to fill your refrigerator and pantry with your personal favorites and staples when you need them. After selecting products through the Instacart app, personal shoppers review your order and do the in-store shopping and delivery for you.

Instacart’s data science team plays a big part in providing this delightful shopping experience. Currently they use transactional data to develop models that predict which products a user will buy again, try for the first time, or add to their cart next during a session. Recently, Instacart open sourced this data - see their blog post on [3 Million Instacart Orders, Open Sourced.](https://tech.instacart.com/3-million-instacart-orders-open-sourced-d40d29ead6f2)

In this competition, Instacart is challenging the Kaggle community to use this anonymized data on customer orders over time to predict which previously purchased products will be in a user’s next order. They’re not only looking for the best model, Instacart’s also looking for [machine learning engineers](https://boards.greenhouse.io/instacart/jobs/590555) to grow their team.

Winners of this competition will receive both a cash prize and a fast track through the recruiting process. For more information about exciting opportunities at Instacart, check out their [careers page here](https://careers.instacart.com/) or e-mail their recruiting team directly at ml.jobs@instacart.com.

Submissions will be evaluated based on their [mean F1 score](https://en.wikipedia.org/wiki/F1_score).

Submission File

For each order*id in the test set, you should predict a space-delimited list of product*ids for that order. If you wish to predict an empty order, you should submit an explicit 'None' value. You may combine 'None' with product\_ids. The spelling of 'None' is case sensitive in the scoring metric. The file should have a header and look like the following:

order\_id,products

17,1 2

34,None

137,1 2 3

etc.

**Data Description**

The dataset for this competition is a relational set of files describing customers' orders over time. The goal of the competition is to predict which products will be in a user's next order. The dataset is anonymized and contains a sample of over 3 million grocery orders from more than 200,000 Instacart users. For each user, we provide between 4 and 100 of their orders, with the sequence of products purchased in each order. We also provide the week and hour of day the order was placed, and a relative measure of time between orders. For more information, see the [blog post](https://tech.instacart.com/3-million-instacart-orders-open-sourced-d40d29ead6f2) accompanying its public release.

## File descriptions

Each entity (customer, product, order, aisle, etc.) has an associated unique id. Most of the files and variable names should be self-explanatory.

### **aisles.csv**

aisle\_id,aisle

1,prepared soups salads

2,specialty cheeses

3,energy granola bars

...

### **departments.csv**

department\_id,department

1,frozen

2,other

3,bakery

...

### **order\_products\_\_\*.csv**

These files specify which products were purchased in each order. order\_products\_\_prior.csv contains previous order contents for all customers. 'reordered' indicates that the customer has a previous order that contains the product. Note that some orders will have no reordered items. You may predict an explicit 'None' value for orders with no reordered items. See the evaluation page for full details.

order\_id,product\_id,add\_to\_cart\_order,reordered

1,49302,1,1

1,11109,2,1

1,10246,3,0

...

### **orders.csv**

This file tells to which set (prior, train, test) an order belongs. You are predicting reordered items only for the test set orders. 'order\_dow' is the day of week.

order\_id,user\_id,eval\_set,order\_number,order\_dow,order\_hour\_of\_day,days\_since\_prior\_order

2539329,1,prior,1,2,08,

2398795,1,prior,2,3,07,15.0

473747,1,prior,3,3,12,21.0

...

### **products.csv**

product\_id,product\_name,aisle\_id,department\_id

1,Chocolate Sandwich Cookies,61,19

2,All-Seasons Salt,104,13

3,Robust Golden Unsweetened Oolong Tea,94,7

...

### **sample\_submission.csv**

order\_id,products

17,39276

34,39276

137,39276

...