Overview Search kaggle ☌ssion Leaderboard Rules Submit Predictions

🏆 Featured Prediction Competition

Instacart Market Basket Analysis

Which products will an Instacart consumer purchase again?

$25,000
Prize Money

🥕 Instacart · 1,545 teams · a month to go

Training Data

📄 aisles.csv.zip

📄 departments.csv.zip

📄 order_products__prio...

📄 order_products__trai...

📄 orders.csv.zip

📄 products.csv.zip

📄 sample_submission.cs...

departments.csv.zip 804 B

⬇ Download

Data Introduction

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 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

Overview   **Data**   Kernels   Discussion   Leaderboard   Rules   Submit Predictions

```
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
```

Overview   **Data**   Kernels   Discussion   Leaderboard   Rules   Submit Predictions

. . .

sample_submission.csv

```
order_id,products
17,39276
34,39276
137,39276
...
```

© 2017 Kaggle Inc

Our Team   Terms   Privacy   Contact/Support