

Exercise:

CF + Non Personalized Recommendation for H&M Challenge

Notebook and Submission are mandatory

Groups of up to **3 people** are allowed

How to submit:

Upload the code at Campus Virtual
+ share the notebook via Kaggle

Deadline: 4 March

The screenshot shows the H&M Personalized Fashion Recommendations competition page. At the top, it features a banner with the title "H&M Personalized Fashion Recommendations" and a prize of "\$50,000". Below the banner, there's a navigation bar with tabs: Overview, Data, Code, Discussion, Leaderboard, Rules, Team, My Submissions, and a Submit Predictions button. The main content area is titled "Data Description" and contains a paragraph explaining the challenge: "For this challenge you are given the purchase history of customers across time, along with supporting metadata. Your challenge is to predict what articles each customer will purchase in the 7-day period immediately after the training data ends. Customer who did not make any purchase during that time are excluded from the scoring." Below this, there's a "Files" section listing the data files: images/, articles.csv, customers.csv, sample_submission.csv, and transactions_train.csv. A note at the bottom states: "NOTE: You must make predictions for all customer_id values found in the sample submission. All customers who made purchases during the test period are scored, regardless of whether they had purchase history in the training data." At the bottom of the page, there's a terminal window showing the command "kaggle competitions download -c h-and-m-personalized-fashion-recommendations" and a "Data Explorer" section showing the file structure of the transactions_train.csv file.

Featured Prediction Competition

H&M Personalized Fashion Recommendations

Provide product recommendations based on previous purchases

\$50,000
Prize Money

H&M Group · 582 teams · 3 months to go (2 months to go until merger deadline)

Overview **Data** Code Discussion Leaderboard Rules Team My Submissions **Submit Predictions** ...

Data Description

For this challenge you are given the purchase history of customers across time, along with supporting metadata. Your challenge is to predict what articles each customer will purchase in the 7-day period immediately after the training data ends. Customer who did not make any purchase during that time are excluded from the scoring.

Files

- images/ - a folder of images corresponding to each `article_id`; images are placed in subfolders starting with the first three digits of the `article_id`; note, not all `article_id` values have a corresponding image.
- articles.csv - detailed metadata for each `article_id` available for purchase
- customers.csv - metadata for each `customer_id` in dataset
- sample_submission.csv - a sample submission file in the correct format
- transactions_train.csv - the training data, consisting of the purchases each customer for each date, as well as additional information. Duplicate rows correspond to multiple purchases of the same item. Your task is to predict the `article_id`s each customer will purchase during the 7-day period **immediately after the training data period**.

NOTE: You must make predictions for all `customer_id` values found in the sample submission. All customers who made purchases during the test period are scored, regardless of whether they had purchase history in the training data.

```
>_ kaggle competitions download -c h-and-m-personalized-fashion-recommendations
```

Data Explorer

34.56 GB

- images
- articles.csv
- customers.csv

transactions_train.csv (3.49 GB)

Detail Compact Column 5 of 5 columns

date	customer_id	article_id	price
------	-------------	------------	-------