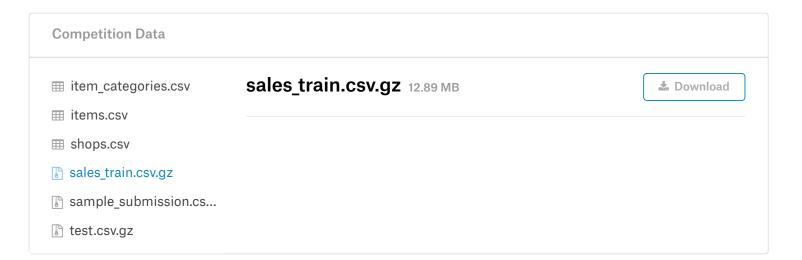**kaggle**   Search kaggle   Q   Competitions   Datasets   Kernels   Discussion   Jobs

🚴 Playground Prediction Competition

Predict Future Sales

Final project for "How to win a data science competition" Coursera course

91 teams · 10 months to go

Overview   **Data**   Kernels   Discussion   Leaderboard   Rules   Join Competition

---

Competition Data

🖥 item_categories.csv
🖥 items.csv
🖥 shops.csv
🗒 sales_train.csv.gz
🗒 sample_submission.cs...
🗒 test.csv.gz

sales_train.csv.gz 12.89 MB   ⬇ Download

```
>_    kaggle competitions download -c competitive-data-science-predict-future-sales
```
📋

Data Description

You are provided with daily historical sales data. The task is to forecast the total amount of products sold in every shop for the test set. Note that the list of shops and products slightly changes every month. Creating a robust model that can handle such situations is part of the challenge.

File descriptions

- **sales_train.csv** - the training set. Daily historical data from January 2013 to October 2015.
- **test.csv** - the test set. You need to forecast the sales for these shops and products for November 2015.
- **sample_submission.csv** - a sample submission file in the correct format.
- **items.csv** - supplemental information about the items/products.
- **item_categories.csv** - supplemental information about the items categories.

- **shops.csv**- supplemental information about the shops.

Data fields

- **ID** - an Id that represents a (Shop, Item) tuple within the test set
- **shop_id** - unique identifier of a shop
- **item_id** - unique identifier of a product
- **item_category_id** - unique identifier of item category
- **item_cnt_day** - number of products sold. You are predicting a monthly amount of this measure
- **item_price** - current price of an item
- **date** - date in format dd/mm/yyyy
- **date_block_num** - a consecutive month number, used for convenience. January 2013 is 0, February 2013 is 1,..., October 2015 is 33
- **item_name** - name of item
- **shop_name** - name of shop
- **item_category_name** - name of item category