



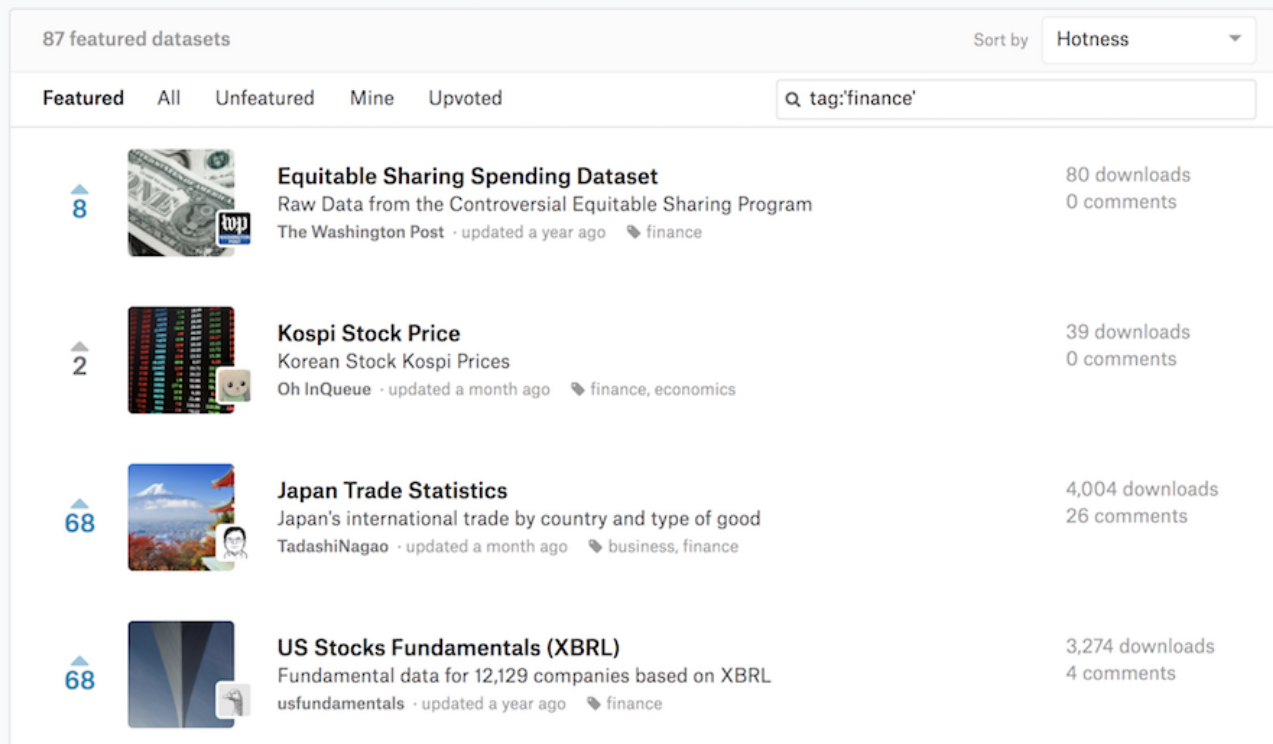
InClass Setup Guide

Creating a Competition

The **competition wizard** is where you will begin creating your competition. The wizard will walk you through the setup of your competition including the different content pages, timeline, evaluation metric, and dataset. Get started by entering a title and subtitle for your competition. After clicking 'Create Competition' you'll land on the Host tab.

Finding a Dataset

To create a competition, you'll first need a clean dataset. We've compiled a **list of machine learning friendly datasets here**. You can also browse **Kaggle's Datasets platform** for data on a specific topic. You can use the search box to find data tagged for a specific topic. Just type: Tag:'[topic]'



The screenshot shows the Kaggle Datasets search interface. At the top, it indicates '87 featured datasets' and a 'Sort by' dropdown menu set to 'Hotness'. Below this is a navigation bar with tabs for 'Featured', 'All', 'Unfeatured', 'Mine', and 'Upvoted'. A search box contains the query 'tag:'finance''. The results are displayed as a list of four datasets, each with a thumbnail, title, description, source, update date, tags, and statistics (downloads and comments).

Rank	Thumbnail	Dataset Title	Description	Source	Updated	Tags	Downloads	Comments
8		Equitable Sharing Spending Dataset	Raw Data from the Controversial Equitable Sharing Program	The Washington Post	updated a year ago	finance	80	0
2		Kospi Stock Price	Korean Stock Kospi Prices	Oh InQueue	updated a month ago	finance, economics	39	0
68		Japan Trade Statistics	Japan's international trade by country and type of good	TadashiNagao	updated a month ago	business, finance	4,004	26
68		US Stocks Fundamentals (XBRL)	Fundamental data for 12,129 companies based on XBRL	usfundamentals	updated a year ago	finance	3,274	4

Setting Up Scoring

You'll need a:

- Solution file
- Sample submission file

For each competition, you will need to prepare and upload a solution file and an sample submission file in csv format. All solutions and submissions should have a unique Id column. The Id column is the only way the scoring system knows which rows of a submission correspond to which rows of the solution.

Cleaned Dataset

Every InClass competition has a public and private leaderboard. In order to manually split your data across these two leaderboards, you need to add a "Usage" column to your dataset. The Usage column should be the last column of the dataset and must have one of two values: Public or Private. If you do not provide a Usage column, rows will automatically be sampled into these buckets. Public rows are used to create the live leaderboard while the competition is running. Private rows are used to determine the final standings and are only viewable by you until the competition has ended.

Id	Prediction1	Prediction2	Usage
1	1	1	Public
2	0	1	Private
3	1	0	Ignored

Here you would map both Prediction1 and Prediction2 from the above dataset to the Expected column, and both columns count equally in the scoring (as if they were one big column).

Solution File

Some metrics let you score multiple columns simultaneously. To do this, map each column of the solution file to the same column of the metric (for most metrics the column is named "Expected"). In this case your solution file might look like this:

Id	Prediction1	Usage
1	1	Public
2	0	Private
3	1	Ignored

Some metrics allow a variable number of predictions so they do not fit into a rectangular CSV format. Most of these metrics are written to accept space-delimited lists. For example:

Id	Prediction1	RetrievedDocuments	Usage
1	1	ABC	Public
2	0	DEFGHI	Private
3	1	J	Ignored

Sample Submission File

The submission file should not contain extraneous information that is not used for scoring. Depending on the metric you choose, you will be asked to map the columns of each file to the columns the metric expects. An example submission file format looks like:

Id	Prediction1
1	0
2	0
3	0

Benchmark Solution


To create a benchmark score for your students to meet or exceed, simply check the box next to the sample submission you'd like to use as a benchmark. You'll then see that score listed as a benchmark on the leaderboard.

Testing a Competition

Once you map the solution and submission files to the metric, you can test submissions in the submission sandbox. This is when you should verify that the scoring is working (e.g. a random submission should have a random score, a perfect submission should have a perfect score, etc.). You may have to revise the format a few times to get it to work, but the system should provide clear error messages in the event something is wrong with a file.

Host Controls

- Settings
- Images
- Privacy
- Manage Hosts
- Evaluation
- Sandbox Submissions**
- All Submissions
- Launch Checklist



Upload File

Name		Date	Public	Private	Benchmark
Ones		4 months ago	0.49523	0.50167	<input checked="" type="checkbox"/>
num_words_800_multiprocess		6 months ago	0.39085	0.39297	<input checked="" type="checkbox"/>
one_feature		8 months ago	0.73283	0.72204	<input checked="" type="checkbox"/>
num_words_800		8 months ago	0.39085	0.39297	<input checked="" type="checkbox"/>
kaggle_submission.csv		8 months ago	0.50476	0.49832	<input type="checkbox"/>
kaggle_submission.csv		8 months ago	0.50476	0.49832	<input type="checkbox"/>
kaggle_submission.csv		8 months ago	0.50476	0.49832	<input type="checkbox"/>

Inviting Students

You can invite students to your competition by sharing the unique URL from the Host > Privacy tab. Anyone with this URL can participate in the competition so make sure it is not circulated beyond your classroom.

Host Controls	
Settings	
Images	
Privacy	Privacy <input type="text" value="Limited"/> <p>Participation of this competition is restricted to those with access to the link below, or to those included in the whitelist below. Note that anyone is still able to view the details of this competition.</p>
Manage Hosts	URL for Sharing <input type="text" value="https://www.kaggle.com/t/5e0c482a7da541a998f73fb6785"/> <input type="button" value="Copy"/> <input type="button" value="Refresh"/> <p>This is a URL that can be shared around and re-used. Anyone who visits this link will be able to participate in the competition.</p>
Evaluation	
Sandbox Submissions	
All Submissions	
Launch Checklist	<input type="checkbox"/> Whitelist users <div style="border: 1px solid #ccc; height: 60px; margin-top: 5px;"></div> <p>Enter a list of user IDs (delimited by anything) to grant access to this competition. Note: any users from this list will be added to any existing users that already have access, and cannot be removed once whitelisted.</p> <p style="text-align: center;"><input type="button" value="Save Competition"/></p>

Launching a Competition

Go to Host > Launch Checklist and confirm that all the boxes are checked green. Once they are, you're good to go! You'll know your competition is live when it says "Competition is active."

Host Controls	
Settings	<input checked="" type="checkbox"/> You have uploaded a solution
Images	<input checked="" type="checkbox"/> You have mapped your solution columns to the metric's expected columns.
Privacy	<input checked="" type="checkbox"/> You have mapped your submission columns to the metric's expected columns.
Manage Hosts	<input checked="" type="checkbox"/> You have made a successful sandbox submission
Evaluation	<input checked="" type="checkbox"/> You have uploaded at least one data file
Sandbox Submissions	<input checked="" type="checkbox"/> The deadline is in the future
All Submissions	<input checked="" type="checkbox"/> You have published all required pages
Launch Checklist	<input checked="" type="checkbox"/> You have no remaining template content
	Your competition is active.