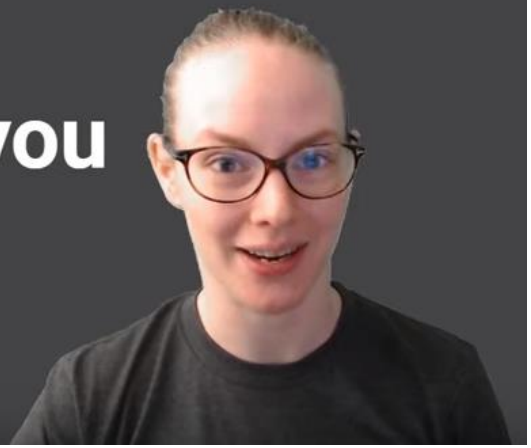
Somebody: “Check out my kernel! It’s doing well on the public LB and I think it’ll still be high ranking after the shakeup.”

You: “..... What on earth are you talking about?”

kaggle



Kaggle

- Home of data science
  - Code + data + forums + competitions + courses + you!
- Made-up word
- Rhymes with “gaggle” (a group of geese)

Gaggle of geese ^^

This is Kerneler.

Kernels

- Hosted coding environment
- Runs in the browser (no downloading)
- Write, run & share data science code

Types:

- **Notebooks**: hosted Jupyter notebooks (text + code)
- **Scripts**: Raw Python, R or R Markdown files (code only)

This is Kerneler.

More Kernels Stuff

These are two cells from a notebook kernel

- **Cell**: in notebooks, cells are editable blocks of text or code
- **Docker**: A docker is basically a whole file system. We have two (one in Python, one in R) with everything your need to run code written on Kaggle locally. You can download them from the Kaggle GitHub.
- **Commit**: Commiting code runs it from top to bottom & saves a static version you can refer back to. (Like a save file in a video game.)

Competitions

A 🌈rainbow🌈 of competitions types:

- Featured: The most competitive competition with the biggest prizes. Often hosted by corporations looking to improve their current approach.
- Analytics: Not supervised machine learning. Many of these are designed to help nonprofits with tricky problems they're facing.
- Research: More theoretical problems based on active areas of research (academic & industrial).
- Getting Started: Designed to be your first or second competition.
- Playground: Especially interesting problems we think you might enjoy trying to solve or fun seasonal competitions.

Competitions

A 🌈rainbow🌈 of competitions types:

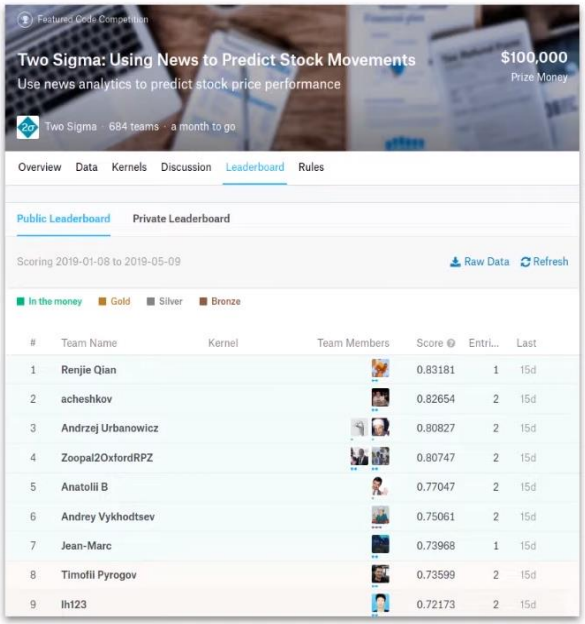
- Featured: The most competitive competition with the biggest prizes. Often hosted by ... current approach. ...ese are ... facing.
- ... reas of research ...
- Getting Started: Designed to be your first or second competition.
- Playground: Especially interesting problems we think you might enjoy trying to solve or fun seasonal competitions.

Only featured (and some research) competitions award **points**. Points are used to determine overall user rankings.

Competitions

- **InClass**: Lets you host your own Kaggle competition. (They don't have to be for an actual class, that's just why is was started.)
- **Leaderboard**: The place where all the scores for entries to a specific competition are shown.
  - **Public**: This is the one that you see during the competition.
  - **Private**: You only see this after the competition ends. Usually scored on a different part of the data.
- **Shakeup:** Changes in ranking when the private leaderboard is shown at the end of a competition. (Usually people move down the ranks due to overfitting.)

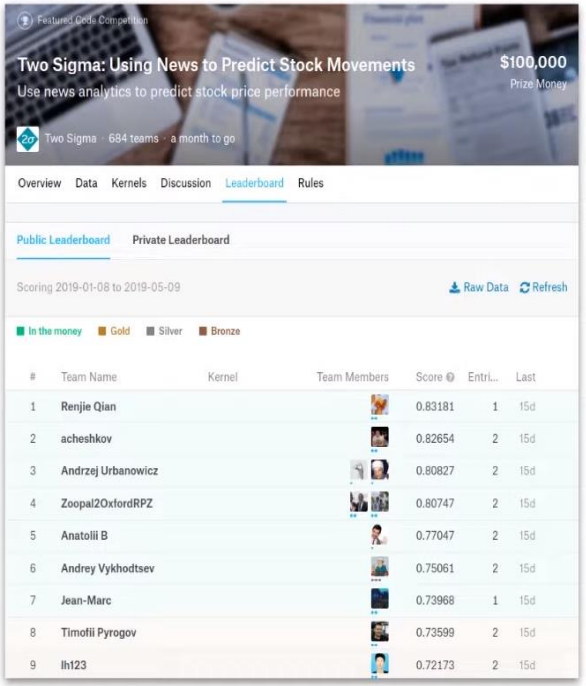
Sample leaderboard for a competition



Competitions

"In the money" means that people at that rank will win money. (If they still have that rank on the private leaderboard after the shakeup.)

Sample leaderboard for a competition



Leakage

- When information about the target variable is accidentally included in the training data
  - All photos of the target class were taken with a specific camera, so you end up actually building a camera type detector
  - You're trying to predict if someone will get an illness, but you don't remove information about whether someone was *treated* for that illness
- Kaggle data scientists work very hard to minimize leakage in competitions

Kaggle Progression System

Novice → Contributor → Expert → Master → Grandmaster

- Once you reach Contributor, your votes count for awarding other people medals!
- Once you reach Expert, you get ranked separately in three categories: **discussion, kernels** and **competitions**
- **Medals**: These are awarded for having your content upvoted or being near/at the top of the leaderboard for competitions.

Kaggle Progression System

Novice → Contributor → Expert → Master → Grandmaster

Out of >3 Million Kagglers there are:

- **Competitions:** 145 Grandmasters, 1220 Masters
- **Kernels**: 11 Grandmasters, 49 Masters
- **Discussion:** 12 Grandmasters, 20 Masters

Counts taken June, 2019

It's pretty hard to become a Master or Grandmaster,
but you can get there with hard work & dedication! 💪💪