

The Challenge

The sinking of the Titanic is one of the most infamous shipwrecks in history.

On April 15, 1912, during her maiden voyage, the widely considered “unsinkable” RMS Titanic sank after colliding with an iceberg. Unfortunately, there weren't enough lifeboats for everyone onboard, resulting in the death of 1502 out of 2224 passengers and crew.

While there was some element of luck involved in surviving, it seems some groups of people were more likely to survive than others.

In this challenge, we ask you to build a predictive model that answers the question: “what sorts of people were more likely to survive?” using passenger data (ie name, age, gender, socio-economic class, etc).

Recommended Tutorial

We highly recommend [Alexis Cook's Titanic Tutorial](#) that walks you through making your very first submission step by step and [this starter notebook](#) to get started.

How Kaggle's Competitions Work

1. Join the Competition

Read about the challenge description, accept the Competition Rules and gain access to the competition dataset.

2. Get to Work

Download the data, build models on it locally or on Kaggle Notebooks (our no-setup, customizable Jupyter Notebooks environment with free GPUs) and generate a prediction file.

3. Make a Submission

Upload your prediction as a submission on Kaggle and receive an accuracy score.

4. Check the Leaderboard

See how your model ranks against other Kagglers on our leaderboard.

5. Improve Your Score

Check out the [discussion forum](#) to find lots of tutorials and insights from other competitors.

Kaggle Lingo Video

You may run into unfamiliar lingo as you dig into the Kaggle discussion forums and public notebooks. Check out Dr. Rachael Tatman's [video on Kaggle Lingo](#) to get up to speed!

What Data Will I Use in This Competition?

In this competition, you'll gain access to two similar datasets that include passenger information like name, age, gender, socio-economic class, etc. One dataset is titled `train.csv` and the other is titled `test.csv`.

`Train.csv` will contain the details of a subset of the passengers on board (891 to be exact) and importantly, will reveal whether they survived or not, also known as the “ground truth”.

The `test.csv` dataset contains similar information but does not disclose the “ground truth” for each passenger. It's your job to predict these outcomes.

Using the patterns you find in the `train.csv` data, predict whether the other 418 passengers on board (found in `test.csv`) survived.

Check out the [“Data” tab](#) to explore the datasets even further. Once you feel you've created a competitive model, submit it to Kaggle to see where your model stands on our leaderboard against other Kagglers.