Create

Home

Competitions

Datasets

Code

Discussions

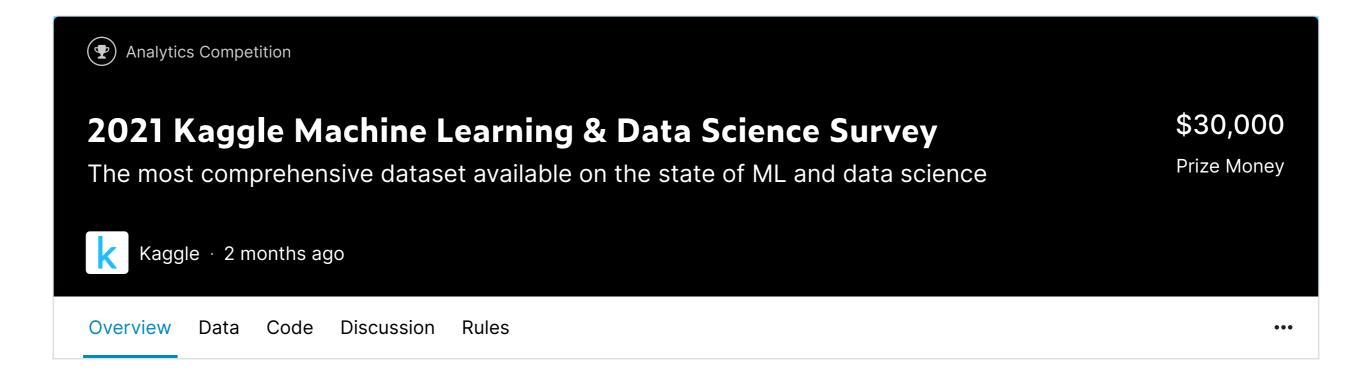
Courses

More

Q Search

Sign In

Register



Overview

Description

Evaluation

Timeline

Prizes

Welcome to Kaggle's annual Machine Learning and Data Science Survey competition! You can read our executive summary here.

This year, as in 2017, 2018, 2019, and 2020 we set out to conduct an industry-wide survey that presents a truly comprehensive view of the state of data science and machine learning. The survey was live from 09/01/2021 to 10/04/2021, and after cleaning the data we finished with 25,973 responses!

There's a lot to explore here. The results include raw numbers about who is working with data, what's happening with machine learning in different industries, and the best ways for new data scientists to break into the field. We've published the data in as raw a format as possible without compromising anonymization, which makes it an unusual example of a survey dataset.

This year Kaggle is once again launching an annual Data Science Survey Challenge, where we will be awarding a prize pool of \$30,000 to notebook authors who tell a rich story about a subset of the data science and machine learning community.

In our fifth year running this survey, we were once again awed by the global, diverse, and dynamic nature of the data science and machine learning industry. This survey data EDA provides an overview of the industry on an aggregate scale, but it also leaves us wanting to know more about the many specific communities comprised within the survey. For that reason, we're inviting the Kaggle community to dive deep into the survey datasets and help us tell the diverse stories of data scientists from around the world.

The challenge objective: tell a data story about a subset of the data science community represented in this survey, through a combination of both narrative text and data exploration. A "story" could be defined any number of ways, and that's deliberate. The challenge is to deeply explore (through data) the impact, priorities, or concerns of a specific group of data science and machine learning practitioners. That group can be defined in the macro (for example: anyone who does most of their coding in Python) or the micro (for example: female data science students studying machine learning in masters programs). This is an opportunity to be creative and tell the story of a community you identify with or are passionate about!

Submissions will be evaluated on the following:

- Composition Is there a clear narrative thread to the story that's articulated and supported by data? The subject should be well defined, well researched, and well supported through the use of data and visualizations.
- Originality Does the reader learn something new through this submission? Or is the reader challenged to think about something in a new way? A great entry will be informative, thought provoking, and fresh all at the same time.
- Documentation Are your code, and notebook, and additional data sources well documented so a