

Project Proposal Template

2021 Kaggle Machine Learning & Data Science Survey

Overview About the data:

This year, Kaggle set out to conduct an industry-wide survey that presents a truly comprehensive view of the state of data science and machine learning and the best ways for new data scientists to break into the field. Survey data EDA provides an overview of the sector on an aggregate scale. The survey questions range from demographic questions, such as gender and higher education level, to questions about programming languages, tools, and machine learning algorithms used. The results include raw numbers about who is working with data and machine learning in different industries.

Data Description:

Data The dataset is provided in .csv format. It contains 369 columns and 25974 rows. Columns are all questions asked in the questionnaire and rows are responses from the Kaggle community. (<https://www.kaggle.com/c/kaggle-survey-2021/overview>)

Questions:

- What's the most popular programming language in 2021?
- What are the 5 countries with the most awareness of data science and artificial intelligence from the 2021 survey?
- What is the gender distribution of data scientists in the Kaggle community?
- What are the salaries for data scientists in the Kaggle community?
- What's the most used of (IDE's) in 2021?
- What are the most used hosted notebook products in 2021? ence?

Tools:

- Pandas for data manipulation.
- numpy summary statistics .
- Matplotlib for plotting.
- Seaborn for data visualization.

