

## 2022 Kaggle DS & ML Survey

### List of Questions and Answer Choices

#### C o n t e x t :

The goal of the  
Kaggle DS & ML survey  
is

to better understand current trends in data  
science and machine learning. This document describes the questions and answer choices that  
were presented to survey respondents.

To find the list of questions and answer choices from previous years (along with the raw data),  
you can visit the following links:

2017

,  
2018

,  
2019

,  
2020

,  
2021

,  
2022

.

Questions with answer choices:

Q1

Welcome to the 2022 Kaggle Machine Learning and Data Science Survey! It should take roughly 10 to  
minutes to complete this survey. Anonymized survey results will be released publicly at the end of the  
year.

Q2

1

What is your age (# years)?

[List of Values]

Q3

What is your gender?

Man

Woman

Nonbinary

Prefer not to say

Prefer to self-describe

Q4

In which country do you currently reside?

[List of Countries]

Q5

Are you currently a student? (high school, university, or graduate)

Yes

No

1

If Q5 < 18 then no additional data is collected

Q6

On which platforms have you begun or completed data science courses? (Select all that apply)

Coursera

edX

Kaggle Learn Courses

DataCamp

Fast.ai

Udacity

Udemy

LinkedIn Learning

Cloud-certification programs (direct from AWS, Azure, GCP, or similar)

University Courses (resulting in a university degree)

None

Other

Q7

What products or platforms did you find to be most helpful when you first started studying data science? (Select all that apply)

University courses

Online courses (Coursera, EdX, etc)

Social media platforms (Reddit, Twitter, etc)

Video platforms (YouTube, Twitch, etc)

Kaggle (notebooks, competitions, etc)

None / I do not study data science

Other

Q8

What is the highest level of formal education that you have attained or plan to attain within the next 2 years?

No formal education past high school

Some college/university study without earning a bachelor's degree

Bachelor's degree

Master's degree

Doctoral degree

Professional doctorate

I prefer not to answer

Q9

2

Have you ever published any academic research (papers, preprints, conference proceedings, etc)?

Yes

No

Q10

3

Did your research make use of machine learning? (select multiple)

Yes, the research made advances related to some novel machine learning method (theoretical research)

Yes, the research made use of machine learning as a tool (applied research)

No

Q11

For how many years have you been writing code and/or programming?

I have never written code

< 1 years

1-2 years  
3-5 years  
5-10 years  
10-20 years  
20+ years

Q12

4

What programming languages do you use on a regular basis? (Select all that apply)

Python

R

SQL

C

C++

Java

Javascript

Julia

Bash

MATLAB

None

Other

C#

4

If Q10 > 0

3

Only asked if Q8 = Yes

2

Only asked if Q7 > Bachelor's

PHP

Go

Q13

5

Which of the following integrated development environments (IDE's) do you use on a regular basis?  
(Select all that apply)

JupyterLab

RStudio / Posit

Visual Studio

Visual Studio Code (VSCode)

PyCharm

Spyder

Notepad++

Sublime Text

Vim, Emacs, or similar

MATLAB

Jupyter Notebook

IntelliJ

None

Other

Q14

6

Do you use any of the following hosted notebook products? (Select all that apply)

Kaggle Notebooks

Colab Notebooks

Azure Notebooks

Code Ocean

IBM Watson Studio

Amazon Sagemaker Studio

Amazon Sagemaker Studio Lab

Amazon EMR Notebooks

Google Cloud Vertex AI Workbench

Hex Workspaces

Noteable Notebooks

Databricks Collaborative Notebooks

Deepnote Notebooks

Gradient Notebooks

None

Other

6

If Q10 > 0

5

If Q10 > 0

Q15

7

Do you use any of the following data visualization libraries on a regular basis? (Select all that apply)

Matplotlib

Seaborn

Plotly / Plotly Express

Ggplot / ggplot2

Shiny

D3.js

Altair

Bokeh

Geoplotlib

Leaflet / Folium

Pygal

Dygraphs

Highcharter

None

Other

Q16

8

For how many years have you used machine learning methods?

I do not use machine learning methods

Under 1 year

1-2 years

2-3 years

3-4 years

4-5 years

5-10 years

10-20 years

20 or more years

8

If Q10 > 0

7

If Q10 > 0

Q17

9

Which of the following machine learning frameworks do you use on a regular basis? (Select all that apply)

Scikit-learn

TensorFlow

Keras

PyTorch

Fast.ai

Xgboost

LightGBM

CatBoost

Caret

Tidymodels

JAX

PyTorch Lightning

Huggingface

None

Other

Q18

10

Which of the following ML algorithms do you use on a regular basis? (Select all that apply):

Linear or Logistic Regression

Decision Trees or Random Forests

Gradient Boosting Machines (xgboost, lightgbm, etc)

Bayesian Approaches

Evolutionary Approaches

Dense Neural Networks (MLPs, etc)

Convolutional Neural Networks

Generative Adversarial Networks

Recurrent Neural Networks

Transformer Networks (BERT, gpt-3, etc)

Autoencoder Networks (DAE, VAE, etc)

Graph Neural Networks

None

Other

10

If Q15 > 0

9

If Q15 > 0

Q19

11

Which categories of computer vision methods do you use on a regular basis? (Select all that apply)

General purpose image/video tools (PIL, cv2, skimage, etc)  
Image segmentation methods (U-Net, Mask R-CNN, etc)  
Object detection methods (YOLOv6, RetinaNet, etc)  
Image classification and other general purpose networks (VGG, Inception, ResNet, ResNeXt, NASNet, EfficientNet, etc)  
Vision transformer networks (ViT, DeiT, BiT, BEiT, Swin, etc)  
Generative Networks (GAN, VAE, etc)

None

Other

Q20

12

Which of the following natural language processing (NLP) methods do you use on a regular basis?  
(Select all that apply)

Word embeddings/vectors (GLoVe, fastText, word2vec)

Encoder-decoder models (seq2seq, vanilla transformers)

Contextualized embeddings (ELMo, CoVe)

Transformer language models (GPT-3, BERT, XLnet, etc)

None

Other

Q21

13

Do you download pre-trained model weights from any of the following services? (Select all that apply)

Tfhub.dev

Pytorch hub

Huggingface models

Timm

Jumpstart

ONNX models

NVIDIA NGC models

Kaggle datasets

Other storage services (i.e. google drive)

I do not download pre-trained model weights on a regular basis

13

If Q15 > 0

12

If Q15 > 0 and Q17 = RNN or Transformer

11

If Q15 > 0 and Q17 = CNN, GAN, or Transformer

Q22

14

Which of the following ML model hubs/repositories do you use most often? (Select all that apply)

»

Tfhub.dev

»

Pytorch hub

»

Huggingface models

»

Timm

»

Jumpstart

»

ONNX models

»

NVIDIA NGC models

»

Kaggle datasets

»

Other storage services (i.e. google drive)

Q23

15

Select the title most similar to your current role (or most recent title if retired):

Data Analyst

(Business, Marketing, Financial, Quantitative,  
etc)

Data Architect

Data Engineer

Data Scientist

Data Administrator

Developer Advocate

Machine Learning/ MLOps Engineer

Manager (Program, Project, Operations, Executive-level, etc)

Research Scientist

Software Engineer

Engineer (non-software)

Statistician

Teacher / professor

Currently not employed

Other

15

If Q5 = No

14

If Q20 selected count > 1

Q24

16

In what industry is your current employer/contract (or your most recent employer if retired)?

Academics/Education

Accounting/Finance

Broadcasting/Communications

Computers/Technology

Energy/Mining

Government/Public Service

Insurance/Risk Assessment

Online Service/Internet-based Services

Marketing/CRM

Manufacturing/Fabrication

Medical/Pharmaceutical

Non-profit/Service

Retail/Sales

Shipping/Transportation

Other

Q25

17

What is the size of the company where you are employed?

0-49 employees

50-249 employees

250-999 employees

1000-9,999 employees

10,000 or more employees

Q26

18

Approximately how many individuals are responsible for data science workloads at your place of business?

0

1-2

3-4

5-9

10-14

15-19

20+

18

If Q5 = No and Q17 != unemployed

17

If Q5 = No and Q17 != unemployed

16

If Q5 = No

Q27

19

Does your current employer incorporate machine learning methods into their business?

We are exploring ML methods (and may one day put a model into production)

We use ML methods for generating insights (but do not put working models into production)

We recently started using ML methods (i.e., models in production for less than 2 years)

We have well established ML methods (i.e., models in production for more than 2 years)

No (we do not use ML methods)

I do not know

Q28

20

Select any activities that make up an important part of your role at work: (Select all that apply)

Analyze and understand data to influence product or business decisions

Build and/or run the data infrastructure that my business uses for storing, analyzing, and operationalizing data

Build prototypes to explore applying machine learning to new areas

Build and/or run a machine learning service that operationally improves my product or workflows

Experimentation and iteration to improve existing ML models

Do research that advances the state of the art of machine learning

None of these activities are an important part of my role at work



Other

Q29

21

What is your current yearly compensation (approximate

\$USD

)?

[List of Values]

21

If Q5 = No and Q17 != unemployed

20

If Q5 = No and Q17 != unemployed

19

If Q5 = No and Q17 != unemployed

Q30

22

Approximately how much money have you spent on machine learning and/or cloud computing services at home or at work in the past 5 years (

approximate

\$USD

)?

\$0 (\$USD)

\$1-\$99

\$100-\$999

\$1000-\$9,999

\$10,000-\$99,999

\$100,000 or more (\$USD)

Q31

23

Which of the following cloud computing platforms do you use? (Select all that apply)

Amazon Web Services (AWS)

Microsoft Azure

Google Cloud Platform (GCP)

IBM Cloud / Red Hat

Oracle Cloud

SAP Cloud

VMware Cloud

Alibaba Cloud

Tencent Cloud

Huawei Cloud

None

Other

23

If Q29 > 0

22

If Q5 = No and Q17 != unemployed

Q32

24

Of the cloud platforms that you are familiar with, which has the best developer experience (most

enjoyable to use)?

»

Amazon Web Services (AWS)

»

Microsoft Azure

»

Google Cloud Platform (GCP)

»

IBM Cloud / Red Hat

»

Oracle Cloud

»

SAP Cloud

»

Salesforce Cloud

»

VMware Cloud

»

Alibaba Cloud

»

Tencent Cloud

»

Huawei Cloud

None were satisfactory

They all had a similarly enjoyable developer experience

Other

Q33

25

Do you use any of the following cloud computing products? (Select all that apply)

Amazon Elastic Compute Cloud (EC2)

Microsoft Azure Virtual Machines

Google Cloud Compute Engine

No / None

Other

Q34

26

Do you use any of the following data storage products? (Select all that apply)

Amazon Simple Storage Service (S3)

Amazon Elastic File System (EFS)

Google Cloud Storage (GCS)

Google Cloud Filestore

Microsoft Azure Blob Storage

Microsoft Azure Files

No / None

Other

26

If Q30 = AWS, Azure, or GCP

25

If Q30 = AWS, Azure, or GCP

24

If Q30 selected count > 0

Q35

27

Do you use any of the following data products (relational databases, data warehouses, data lakes, or similar)? (Select all that apply)

MySQL

PostgreSQL

SQLite

Oracle Database

MongoDB

Snowflake

IBM Db2

Microsoft SQL Server

Microsoft Azure SQL Database

Amazon Redshift

Amazon RDS

Amazon DynamoDB

Google Cloud BigQuery

Google Cloud SQL

None

Other

Q36

28

Do you use any of the following business intelligence tools? (Select all that apply)

Amazon QuickSight

Microsoft Power BI

Google Data Studio

Looker

Tableau

Qlik Sense

Domo

TIBCO Spotfire

Alteryx

Sisense

SAP Analytics Cloud

Microsoft Azure Synapse

Microstrategy

None

Other

28

If Q29 > 0

27

If Q29 > 0

Q37

29

Do you use any of the following managed machine learning products? (Select all that apply)

Amazon SageMaker

Azure Machine Learning Studio

Google Cloud Vertex AI  
DataRobot  
Databricks  
Dataiku  
Alteryx  
Rapidminer  
C3.ai  
Domino Data Lab  
H2O AI Cloud  
No / None  
Other

Q38

30

Do you use any of the following automated machine learning tools? (Select all that apply)

Google Cloud AutoML  
H2O Driverless AI  
Databricks AutoML  
DataRobot AutoML  
Amazon Sagemaker Autopilot  
Azure Automated Machine Learning  
No / None  
Other

Other

30

If Q29 > 0 and Q15 > 0

29

If Q29 > 0 and Q15 > 0

Q39

31

Do you use any of the following products to serve your machine learning models? (Select all that apply)

TensorFlow Extended (TFX)  
TorchServe  
ONNX Runtime  
Triton Inference Server  
OpenVINO Model Server  
KServe  
BentoML  
Multi Model Server (MMS)  
Seldon Core

MLflow

Other

None

Q40

32

Do you use any tools to help monitor your machine learning models and/or experiments? (Select all that apply)

Neptune.ai  
Weights & Biases  
Comet.ml

TensorBoard

Guild.ai

ClearML

MLflow

Aporia

Evidently AI

Arize

WhyLabs

Fiddler

DVC

No / None

Other

32

If Q29 > 0 and Q15 > 0

31

If Q29 > 0 and Q15 > 0

Q41

33

Do you use any of the following responsible or ethical AI products in your machine learning practices? (Select all that apply)

Google Responsible AI Toolkit

(LIT, What-If Tools,

Fairness Indicator, TensorFlow Data

Validation, TensorFlow Privacy, etc.)

Microsoft Responsible AI Toolbox

(Fairlearn, Counterfit,

InterpretML, SmartNoise, etc.)

IBM AI Ethics tools

(AI Fairness 360, Adversarial

Robustness Toolbox, AI Explainability 360, etc.)

Amazon AI Ethics Tools

(Clarify, A2I, etc)

The LinkedIn Fairness Toolkit

(LiFT)

Audit-AI

Aequitas

None

Other

Q42

34

Do you use any of the following types of specialized hardware when training machine learning models? (Select all that apply)

GPUs

TPUs

IPUs

WSEs

RDUes

Trainium Chips

Inferentia Chips

None

Other

Q43

35

Approximately how many times have you used a TPU (tensor processing unit)?

Never

Once

2-5 times

6-25 times

More than 25 times

35

If  $Q29 > 0$  and  $Q15 > 0$

34

If  $Q29 > 0$  and  $Q15 > 0$

33

If  $Q29 > 0$  and  $Q15 > 0$

Q44

Who/what are your favorite media sources that report on data science topics? (Select all that apply)

Twitter (data science influencers)

Email newsletters (Data Elixir, O'Reilly Data & AI, etc)

Reddit (r/machinelearning, etc)

Kaggle (notebooks, forums, etc)

Course Forums (forums.fast.ai, Coursera forums, etc)

YouTube (Kaggle YouTube, Cloud AI Adventures, etc)

Podcasts (Chai Time Data Science, O'Reilly Data Show, etc)

Blogs (Towards Data Science, Analytics Vidhya, etc)

Journal Publications (peer-reviewed journals, conference proceedings, etc)

Slack Communities (ods.ai, kagglenoobs, etc)

None

Other