≡ **kaggle**

🔍 Search   **Sign In**   **Register**

+ Create

⊘ Home

♔ Competitions

🎲 Datasets

<> Code

▤ Discussions

🎓 Courses

⌄ More

ANDRADA OLTEANU · UPDATED 5 MONTHS AGO   ▲ 23   **New Notebook**   ⬇ **Download (19 MB)**   ⊘   ⋮

Kaggle Data Science Survey 2017-2021

Merged datasets from 2017 to 2021

Data   Code (7)   Discussion (0)   Metadata

About Dataset

Context

I have created this dataset for an easier way to analyse the progression of answers from the respondents that are participating each year in the very famous Data Science Kaggle Survey.

The sources of the present data are:

- 2017: https://www.kaggle.com/kaggle/kaggle-survey-2017
- 2018: https://www.kaggle.com/kaggle/kaggle-survey-2018
- 2019: https://www.kaggle.com/c/kaggle-survey-2019/data
- 2020: https://www.kaggle.com/c/kaggle-survey-2020/data

Usability ⓘ
9.41

License
Other (specified in description)

Expected update frequency
Never

🗗 View Active Events

https://www.kaggle.com/datasets/andradaolteanu/kaggle-data-science-survey-20172021

kaggle

Q Search

Sign In   Register

+ Create

Home

Competitions

Datasets

<> Code

Discussions

Courses

∨ More

Kaggle Data Science Survey 2017-2021

Data   Code (7)   Discussion (0)   Metadata

▲ 23   New Notebook   ⬇ Download (19 MB)   ⋮

Computer Science   Tabular Data   Exploratory Data Analysis   Survey Analysis

Data Explorer

Version 41 (109.19 MB)

kaggle_survey_2017_2021.csv (97.42 MB)

⬇ ⛶ >

▸ 📁 images

🔲 kaggle_survey_2017_202...

Detail   Compact   Column

10 of 293 columns ∨

📄 style.css

About this file

The aggregated data from all Kaggle Surveys starting 2017 to 2021.

▲ -		▲ Time from Start t...		▲ Q1		▲ Q2		▲ Q3		▲
Year of the answer		Time spent answering the survey		Age of respondents		Gender of respondents		Residence of respondents		Ed
2021	24%	[null]	16%	25-29	22%	Male	46%	India	24%	Ma
2018	22%	394	0%	22-24	18%	Man	34%	United States of A...	12%	Ba
Other (56470)	53%	Other (89479)	84%	Other (62892)	59%	Other (20737)	20%	Other (68422)	64%	Ot
Year		Duration (in seconds)		What is your age (# years)?		What is your gender? - Selected Choice		In which country do you currently reside?		Wh le ed ha pl wi
2021		910		50-54		Man		India		Ba

View Active Events

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	-	Time from	Q1	Q2	Q3	Q4	Q5	Q6	Q7_Part_1	Q7_Part_2	Q7_Part_3	Q7_Part_4	Q7_Part_5	Q7_Part_6	Q7_Part_7	Q7_Part_8	Q7_Part_9	Q7_Part_1	Q7_Part_1	Q7_Part_1	Q7_OTHEF	Q8	Q9_Part_1 Q9
2	Year	Duration (i	What is yo	What is yo	In which c	What is th	Select the	For how m	What prog	What prog	What prog	What prog	What prog	What prog	What prog	What prog	What prog	What prog	What prog	What prog	What prog	What prog	Which of t W
3	2021	910	50-54	Man	India	Bachelorâ	Other	5-10 years	Python	R												Python	
4	2021	784	50-54	Man	Indonesia	Masterâ€™	Program/P	20+ years			SQL	C	C++	Java								Python	
5	2021	924	22-24	Man	Pakistan	Masterâ€™	Software E	1-3 years	Python				C++	Java								Python	
6	2021	575	45-49	Man	Mexico	Doctoral d	Research S	20+ years	Python													Python	
7	2021	781	45-49	Man	India	Doctoral d	Other	< 1 years	Python			C							MATLAB			Python	
8	2021	1020	25-29	Woman	India	I prefer no	Currently r	< 1 years	Python													Python	Jupyter (Jupyt
9	2021	141	18-21	Woman	India	Some colle	Student	1-3 years					C++	Java	Javascript							Python	
10	2021	484	30-34	Man	India	Bachelorâ	Data Scien	5-10 years	Python													Python	
11	2021	1744	22-24	Man	Russia	Bachelorâ	Currently r	3-5 years	Python		SQL											Python	
12	2021	655	30-34	Man	Turkey	I prefer no	Other	1-3 years	Python		SQL											SQL	
13	2021	1777	40-44	Man	Australia	Doctoral d	Other	1-3 years	Python	R	SQL										R	RS
14	2021	3081	18-21	Woman	India	Masterâ€™	Student	< 1 years	Python	R			C++									R	RS
15	2021	1922	18-21	Woman	India	Masterâ€™	Student	< 1 years	Python									MATLAB			Python	RS
16	2021	852	45-49	Man	Nigeria	Masterâ€™	Program/P	5-10 years	Python		SQL											Python	
17	2021	838	22-24	Man	Nigeria	Bachelorâ	Other	< 1 years	Python													Python	
18	2021	563	35-39	Man	Greece	Doctoral d	Research S	10-20 year	Python			C	C++						MATLAB			Python	
19	2021	1315	50-54	Man	Belgium	Bachelorâ	Data Analy	20+ years	Python		SQL											Python	Jupyter (Jupyt
20	2021	479	18-21	Man	Pakistan	Bachelorâ	Data Scien	1-3 years	Python		SQL											Python	Jupyter (Jupyt
21	2021	249	22-24	Man	Japan	Masterâ€™	Software E	3-5 years	Python		SQL	C		Java	Javascript							Python	Jupyter (Jupyt
22	2021	650	30-34	Man	Egypt	Bachelorâ	Other	< 1 years											None			R	
23	2021	1461	70+	Man	Singapore	Bachelorâ	Other	< 1 years	Python													Python	Jupyter (Jupyt
24	2021	551	25-29	Woman	Turkey	Bachelorâ	Data Scien	3-5 years	Python	R	SQL		C++									R	Jupyter (Ju RS
25	2021	258	30-34	Man	Indonesia	Masterâ€™	Student	1-3 years		R												R	RS
26	2021	773	35-39	Man	Brazil	Masterâ€™	Machine L	20+ years	Python		SQL		C++									SQL	
27	2021	630	18-21	Man	India	Bachelorâ	Student	1-3 years	Python	R	SQL	C	C++						MATLAB			Python	RS

kaggle_survey_2017_2021

	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN
1	Q10_Part_	Q10_Part_	Q10_Part_	Q10_OTHE	Q11	Q12new	Q12_Part_	Q12_Part_	Q12_Part_	Q12_OTHE	Q13	Q14_Part_	Q14_Part_	Q14_Part_	Q14_Part_	Q14_Part_	Q14_Part_	Q14_Part_	Q14_Part_	Q14_Part_	Q14_Part_	Q14_OTHE	Q1
2	Which of t	Which of t	Which of t	Which of t	What type	Which type	Which type	Which type	Which type	Which type	Approxima	What data	What data	What data	What data	What data	What data	What data	What data	What data	What data	What data	Fo
3					A laptop			Google Cloud TPUs			2-5 times	Matplotlit	Seaborn		Ggplot / g	Shiny					Leaflet / Folium		5-
4					A cloud computing platform (AWS, Azure, G				None		Never	Matplotlib											Ur
5					A laptop					Other	Never	Matplotlib											I d
6					A cloud computing pl	NVIDIA GPUs					More than	Matplotlib											5-
7	Google Cloud Datalab				A cloud computing platform (AWS, Azure, G				None		Never	Matplotlit	Seaborn		Ggplot / ggplot2								10
8					A laptop			Google Cloud TPUs			2-5 times	Matplotlit	Seaborn		Ggplot / ggplot2								Ur
9	ud Notebooks (AI Platform / Vertex AI)				A laptop				None		Never									Geoplotlib			Ur
10					A personal computer / desktop			Google Cloud TPUs			2-5 times	Matplotlib		Plotly / Plotly Express									2-
11					A cloud computing platform (AWS, Azure, G				None		Never	Matplotlib											Ur
12					A laptop				None		Never	Matplotlit	Seaborn										Ur
13		None			A personal computer / desktop				None		Never	Matplotlit	Seaborn		Ggplot / ggplot2								I d
14					A laptop				None		Never	Matplotlit	Seaborn		Ggplot / ggplot2								Ur
15		None			A laptop																		
16					A laptop				None		Never	Matplotlit	Seaborn										1-
17		None			A laptop		NVIDIA GPUs				Once			Plotly / Plotly Express									Ur
18					A laptop		NVIDIA GPUs				Never	Matplotlib											4-
19			Databricks Collaborative Noteb		A laptop				None		Never	Matplotlit	Seaborn	Plotly / Plotly Express									1-
20					A laptop		NVIDIA GPUs				Never	Matplotlit	Seaborn		Ggplot / ggplot2								Ur
21					A laptop				None		Never	Matplotlib											2-
22		None			A laptop				None		Never											None	I d
23		None			A personal computer / desktop				None		Never	Matplotlit	Seaborn										I d
24					A personal computer		NVIDIA GPUs				Once	Matplotlit	Seaborn		Ggplot / ggplot2								Ur
25		None			A laptop				None		Never				Ggplot / ggplot2								I d
26					A deep learning works	NVIDIA GPUs					Never	Matplotlit	Seaborn									Other	5-
27	Google Cloud Datalab				A personal computer / desktop			Google Cloud TPUs			More than	Matplotlit	Seaborn	Plotly / Pl	Ggplot / g	Shiny							1-

	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN	BO	BP	BQ	BR	BS	BT	BU	BV	BW	BX	BY
13589	Matplotlib												Under 1 year					Fast.ai					
13590	Matplotlib			Ggplot / ggplot2									Under 1 ye	Scikit-lear	TensorFlow		PyTorch						
13591	Matplotli	Seaborn											3-4 years	Scikit-lear	TensorFlc	Keras	PyTorch			Xgboost	LightGBM	CatBoost	
13592			Plotly / Plotly Express										Under 1 ye	Scikit-lear	TensorFlc	Keras	PyTorch			Xgboost			
13593											None		I do not use machine learning methods										
13594	Matplotli	Seaborn											Under 1 ye	Scikit-learn			PyTorch						
13595											None		I do not use machine learning methods										
13596	Matplotlib												Under 1 year				PyTorch						
13597				Ggplot / ggplot2								Other	5-10 years										
13598	Matplotli	Seaborn	Plotly / Pl	Ggplot / g	Shiny	D3 js	Altair	Bokeh	Geoplotli	Leaflet / Folium			Under 1 year		TensorFlc	Keras	PyTorch	Fast.ai		Xgboost	LightGBM	CatBoost	Prophet
13599	Matplotli	Seaborn											Under 1 ye	Scikit-lear	TensorFlow								
13600											None		Under 1 year										
13601	Matplotlib												Under 1 ye	Scikit-learn									
13602	Matplotlib												I do not use machine learning methods										
13603	Matplotli	Seaborn	Plotly / Plotly Express										5-10 years	Scikit-lear	TensorFlow								
13604	Matplotli	Seaborn	Plotly / Plotly Express										1-2 years	Scikit-lear	TensorFlc	Keras				Xgboost	LightGBM		
13605	Matplotli	Seaborn						Bokeh					5-10 years	Scikit-learn									
13606	Matplotli	Seaborn	Plotly / Plotly Express										Under 1 ye	Scikit-learn				Fast.ai					
13607	Matplotlib		Plotly / Plotly Express						Geoplotlib				Under 1 year		TensorFlc	Keras	PyTorch						
13608	Matplotlib												1-2 years			Keras	PyTorch						
13609				Ggplot / ggplot2									Under 1 year										
13610	Matplotlib												Under 1 year										
13611											None		2-3 years		TensorFlow		PyTorch						
13612	Matplotli	Seaborn	Plotly / Plotly Express									Other	1-2 years	Scikit-lear	TensorFlc	Keras							
13613	Matplotlib												Under 1 ye	Scikit-lear	TensorFlc	Keras							
13614	Matplotli	Seaborn											5-10 years	Scikit-learn			PyTorch	Fast.ai		Xgboost			
13615											None		I do not use machine learning methods										

Row	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW
14333																				Other	A laptop	
14334					Other		Kaggle No	Colab Not	Azure Notebooks						Amazon Sagemaker Studio Notebooks						A personal computer	
14335		Notepad++							Colab Notebooks												A cloud computing pla	
14336			Sublime Text																None	A personal computer	
14337								Kaggle Notebooks				Binder / JupyterHub								A deep learning works	
14338		Notepad	Sublime Text					Kaggle Notebooks					IBM Watson Studio			Google Cloud Datalab			A laptop	
14339		Notepad	Sublime Text	MATLAB				Kaggle Notebooks												A laptop	
14340			Sublime Text																Other	A laptop	
14341																			Other	A laptop	
14342								Colab Notebooks						Amazon Sagemaker Studio Notebooks			Databricks Collaborative Noteb	Other	
14343		Notepad++		Vim / Emacs			Other	Kaggle No	Colab Notebooks					Amazon S	Amazon EMR Notebooks					A personal computer	
14344		Notepad++					Other									None			A cloud computing pla	
14345	Spyder				MATLAB				Azure Notebooks							Databricks Collaborative Noteb	A personal computer	
14346								Kaggle Notebooks											A laptop	
14347	Spyder							Kaggle No	Colab Notebooks										A personal computer	
14348	Spyder																		Other	A laptop	
14349				Vim / Emacs				Kaggle No	Colab Notebooks							Databricks Collaborative Noteb	A laptop	
14350									Colab Notebooks										A personal computer	
14351									Colab Notebooks										None	A personal computer	
14352									Colab Notebooks										A laptop	
14353								Kaggle Notebooks								Google Cloud Notebooks (AI Platform / Vertex AI)	A laptop	
14354							Other									None			A laptop	
14355								Kaggle Notebooks											A laptop	
14356		Notepad++														None			A laptop	
14357			Sublime Text													None			A laptop	
14358				Vim / Emacs												None			A laptop	
14359	Spyder															None			A laptop	

Tab 1

Participated Countries vs Maximum Respondents

Tab 2

Gender Distribution in India

Tab 3

Participating Countries World Map

Q3 (Count)

43   25,192

© Mapbox © OpenStreetMap

Tab 4

Age Distribution in India

Q1

● 18-21   ● 22-24   ● 25-29   ● 30-34   ● 35-39   ● 40-44   ● 45-49   ● 50-54   ● 55-59   ● 60-69   ● 70+

Tab 5

Current Role/Title in the Workplace

Q5 (Count)

Tab 6

Experience in Coding

Q6

Category	
3-5 years	
1-2 years	
< 1 years	
5-10 years	
1-3 years	
10-20 years	
< 1 year	
20+ years	
1 to 2 years	
3 to 5 years	
I have never written code	
Less than a year	
More than 10 years	
6 to 10 years	
I have never written code but I want to learn	
I don't write code to analyze data	
20-30 years	
30-40 years	
40+ years	
I have never written code and I do not want to learn	
For how many years have you been writing code and/or progr...	

Average: 4418

Q6 (Count)

0   2,000   4,000   6,000   8,000   10,000   12,000   14,000   16,000

Tab 7

Programming Language used on a Daily Basis

Tab 8

Recommended Programming Language

Q8 (Count)

| 4 | 66,892 |

Q8

✶ Bash　✶ C　✶ C++　✶ C/C++/C#　✶ F#　✶ Go　✶ Haskell　✶ Java　✶ Javascript　✶ Julia
✶ MATLAB　✶ Matlab　✶ Other　✶ Python　✶ R　✶ SAS　✶ SQL　✶ Scala　✶ Stata　✶ Swift
✶ TypeScript　✶ VBA

Tab 9

IDE used on a Regular Basis

Tab 10

Usage of Specialised Hardware on a Regular Basis

Q12_
- ● Google Cloud TPUs ● NVIDIA GPUs

Google Cloud TPUs | 20.0%

NVIDIA GPUs | 80.0%

Tab 11

Usage of Data Visualizations Libraries/Tools on a Regular Basis

Q14_

≡ Altair   ≡ Bokeh   ≡ D3 js   ≡ Geoplotlib   ≡ Leaflet / Folium

≡ Matplotlib   ≡ Plotly / Plotly Express   ≡ Seaborn

Tab 12

Usage of ML Framework on a Regular Basis

Tab 13

Usage of ML Algorithms on a Regular Basis

Q17_

▶ Bayesian Approaches   ▶ Convolutional Neural Netwo...   ▶ Decision Trees or Random F...   ▶ Dense Neural Networks (ML...

▶ Evolutionary Approaches   ▶ Generative Adversarial Netw...   ▶ Gradient Boosting Machine...   ▶ Linear or Logistic Regression

▶ Recurrent Neural Networks   ▶ Transformer Networks (BER...

Tab 14

Cloud Computing Platforms used on a Regular Basis

Q27_ (Count)

7   34

Q27_

- 🟢 Alibaba Cloud
- 🔵 Amazon Web Services (AWS)
- 🔴 Google Cloud Platform (GCP)
- 🟠 IBM Cloud / Red Hat
- 🟢 Microsoft Azure
- 🟣 None
- 🔵 Oracle Cloud
- 🟤 SAP Cloud
- 🟣 Salesforce Cloud
- 🟣 Tencent Cloud
- 🔵 VMware Cloud

Oracle Cloud

None IBM Cloud / Red Hat

Amazon Web Services (AWS)

Google Cloud Platform (GCP)

SAP Cloud   Microsoft Azure   Tencent Cloud   Salesforce Cloud

VMware Cloud   Alibaba Cloud

Tab 15

Data Science Course Platform

Q40_

- 🟡 Coursera
- ⚫ DataCamp
- 🟢 Fast.ai
- 🔵 Kaggle Learn Courses
- 🔵 LinkedIn Learning
- 🟢 Udacity
- 🔵 Udemy
- 🔴 University Courses (resulting...
- 🟣 edX