Package 'DSjobtracker'

July 21, 2025

Type Package

Title What Skills and Qualifications are Required for Data Science
Related Jobs?

Version 2.0.0

Maintainer Thiyanga S. Talagala <ttalagala@sjp.ac.lk>

Description Dataset containing information about job listings for data science job roles.

License CC BY 4.0

URL <https://github.com/thiyangt/DSjobtracker>

Encoding UTF-8

LazyData true

LazyDataCompression xz

RoxygenNote 7.2.3

Depends R (>= 3.5.0)

Suggests knitr, rmarkdown, tibble, tidyr, ggplot2, dplyr, magrittr,
testthat, wordcloud2, forcats, viridis

VignetteBuilder knitr

NeedsCompilation no

Author Thiyanga S. Talagala [aut, cre] (ORCID:
<https://orcid.org/0000-0002-0656-9789>),
Janith C. Wanniarachchi [aut],
Hansani Piyumika [ctb],
HLS Perera [ctb],
S Dissanayake [ctb],
MB Senanayake [ctb],
T Wickramrathne [ctb]

Repository CRAN

Date/Publication 2023-12-09 07:40:02 UTC

Contents

Index   **[19](#)**

DSraw   *Data Scientists/Data Analyst/ Statistician Job Tracker*

Description

Job advertisements

Usage

```
DSraw
```

Format

A data frame with 551 rows and 152 variables

ID row id

Consultant Name of the consultant

DateRetrieved Date of Data Retrieved

DatePublished Published Date of the Advertisement

Job_title Name of the job category

Company Name of the Company

R If R is required -> 1 ,If not mentioned -> 0

SAS If SAS is required -> 1 , If not mentioned -> 0

SPSS If SPSS is required -> 1 , If not mentioned -> 0

Python If Python is required -> 1 , If not mentioned -> 0

MAtlab If Matlab is required -> 1 , If not mentioned -> 0

Scala If Scala is required -> 1 , If not mentioned -> 0

C# If C# is required -> 1 , If not mentioned -> 0

MS Word If knowledge in MS Word is required -> 1 , If not mentioned -> 0

Ms Excel If knowledge in MS Excel is required -> 1 , If not mentioned -> 0

OLE/DB If knowledge in OLE/DB is required -> 1 , If not mentioned -> 0

Ms Access If Ms Access is required -> 1 , If not mentioned -> 0

Ms PowerPoint If knowledge in Ms Powerpoint is required -> 1 , If not mentioned -> 0

Spreadsheets If knowledge in Spreadsheets is required -> 1 , If not mentioned -> 0

Data_visualization If knowledge inData Visualization is required -> 1 , If not mentioned -> 0

Presentation_Skills If Presentation Skills are required -> 1 , If not mentioned -> 0

Communication If Communication skills are required -> 1 , If not mentioned -> 0

BigData If knowledge in Big Data analysis is required -> 1 , If not mentioned -> 0

Data_warehouse If knowledge in Data Warehouse is required -> 1 , If not mentioned -> 0

cloud_storage If knowledge in Cloud Storage is required -> 1 , If not mentioned -> 0

Google_Cloud If knowledge in Google Cloud is required -> 1 , If not mentioned -> 0

AWS If knowledge in AWS is required -> 1 , If not mentioned -> 0

Machine_Learning If knowledge in Machine Learning is required -> 1 , If not mentioned -> 0

Deep Learning If knowledge in Deep Learning is required -> 1 , If not mentioned -> 0

Computer_vision If knowledge in Computer Vision is required -> 1 , If not #' mentioned -> 0

Java If Java is required -> 1 , If not mentioned -> 0

C++ If C++ is required -> 1 , If not mentioned -> 0

C If C is required -> 1 , If not mentioned -> 0

Linux/Unix If knowledge in Linux/Unix is required -> 1 , If not mentioned -> 0

SQL If SQL is required -> 1 , If not mentioned -> 0

NoSQL If NoSQL is required -> 1 , If not mentioned -> 0

RDBMS If knowledge in RDBMS is required -> 1 , If not mentioned -> 0

Oracle If knowledge in Oracle is required -> 1 , If not mentioned -> 0

MySQL If MYSQL is required -> 1 , If not mentioned -> 0

PHP If PHP is required -> 1 , If not mentioned -> 0

Flash_Actionscript If knowledge in Flash Action Script is required -> 1 , If not mentioned -> 0

SPL If knowledge in SPL is required -> 1 , If not mentioned -> 0

web_design_and_development_tools If knowledge in Web Design and Development Tools is required -> 1 , If not mentioned -> 0

Wordpress If knowledge in Wordpress is required -> 1 , If not mentioned -> 0

AI If Artificial Intelligence is required -> 1 , If not mentioned -> 0

Natural_Language_Processing(NLP) If knowledge in NLP is required -> 1 , If not mentioned -> 0

Microsoft Power BI If knowledge in Microsoft Power BI is required -> 1 , If not mentioned -> 0

Google_Analytics If knowledge in Google Analytics is required -> 1 , If not mentioned -> 0

graphics_and_design_skills If Graphic and Design Skills are required -> 1 , If not mentioned -> 0

Data_marketing If Data Marketing abillity is required -> 1 , If not mentioned -> 0

SEO If knowledge in SEO is required -> 1 , If not mentioned -> 0

Content_Management If knowledge in Content Management is required -> 1 , If not mentioned -> 0

Tableau If knowledge in Tableau is required -> 1 , If not mentioned -> 0

D3 If knowledge in D3 is required -> 1 , If not mentioned -> 0

Alteryx If knowledge in Alteryx is required -> 1 , If not mentioned -> 0

KNIME If knowledge in KNIME is required -> 1 , If not mentioned -> 0

Spotfire If knowledge in Spotfire is required -> 1 , If not mentioned -> 0

Spark If knowledge in Spark is required -> 1 , If not mentioned -> 0

S3 If knowledge in S3 is required -> 1 , If not mentioned -> 0

Redshift If knowledge in Redshift is required -> 1 , If not mentioned -> 0

DigitalOcean If knowledge in Digital Ocean is required -> 1 , If not mentioned -> 0

Javascript If Java Script is required -> 1 , If not mentioned -> 0

Kafka If knowledge in Kafka is required -> 1 , If not mentioned -> 0

Storm If knowledge in Storm is required -> 1 , If not mentioned -> 0

Bash If knowledge in Bash is required -> 1 , If not mentioned -> 0

Hadoop If knowledge in Hadoop is required -> 1 , If not mentioned -> 0

Data_Pipelines If knowledge in Data Pipelines is required -> 1 , If not mentioned -> 0

MPP_Platforms If MPP Platforms is required ->1,If not mentioned-0

Qlik If Qlik is required ->1,If not mentioned ->0

Pig If Pig is required ->1,If not mentioned ->0

Hive If Hive is required ->1,If not mentioned ->0

Tensorflow If Tensorflow is required ->1,If not mentioned ->0

Map/Reduce If Map/Reduce is required ->1,If not mentioned ->0

Impala If Impala is required ->1,If not mentioned ->0

Solr If Sloris required ->1,If not mentioned ->0

Teradata If Teradata is required ->1,If not mentioned ->0

MongoDB If MonoDB is required ->1,If not mentioned ->0

Elasticsearch If Elasticsearch is required ->1,If not mentioned ->0

YOLO If YOLO is required-1 ,If not mentioned-0

agile execution If agile execution is required->1 ,If not mentioned->0

Data_management If the knowledge in data management is required->1 ,If not mentioned->0

pyspark If pyspark is required->1 ,If not mentioned->0

Data_mining If the knowledge in data mining is required->1 ,If not mentioned->0

Data_science If the knowledge in data science is required->1 ,If not mentioned->0

Web_Analytic_tools If the knowledge in Web Analytic tools is required->1 ,If not mentioned->0

IOT If IOT is required->1 ,If not mentioned->0

Numerical_Analysis If the knowledge in Numerical Analysis is required->1 ,If not mentioned->0

Economic If the knowledge in Economic is required->1 ,If not mentioned->0

Finance_Knowledge If Finance_Knowledge is required->1 ,If not mentioned->0

Investment_Knowledge If Investment Knowledge is required->1 ,If not mentioned->0

Problem_Solving If the ability of Problem Solving is required->1 ,If not mentioned->0

Korean_language If the ability of speaking Korean language is required->1 ,If not mentioned->0

Bash\Linux Scripting If Bash\ Linux Scripting is required->1 ,If not mentioned->0

Knowledge_in Required knowledge to do a particular job ,If not mentioned->NA

Experience Minimum experience required for a particular job

City City where the company is located in

Location Country where the company is located in

Educational_qualifications Required educational qualifications

Salary Amount of salary

Team_Handling If the ability of Team Handling is required-1 ,If not mentioned-0

Debtor_reconcilation If the ability of Debtor reconciliation is required-1 ,If not mentioned-0

Payroll_management If the ability of Payroll management is required-1 ,If not mentioned-0

Bayesian If Bayesian knowledge is required-1 ,If not mentioned-0

Optimization If Optimization knowledge is required-1 ,If not mentioned-0

Bahasa Malaysia If Bahasa Malaysia is required-1 ,If not mentioned-0

English proficiency If English proficiency is required-1 ,If not mentioned-0

URL Web address of a particular job advertisement

Search_Term web search term of a particular job advertisement

X109 Columns with null values

X110 Columns with null values

X111 Columns with null values

X112 Columns with null values

X113 Columns with null values

X114 Columns with null values

X115 Columns with null values

X116 Columns with null values

X117 Columns with null values

X118 Columns with null values

X119 Columns with null values

X120 Columns with null values

X121 Columns with null values

X122 Columns with null values

X123 Columns with null values

X124 Columns with null values

X125 Columns with null values

X126 Columns with null values

X127 Columns with null values

X128 Columns with null values

X129 Columns with null values

X130 Columns with null values

X131 Columns with null values

X132 Columns with null values

X133 Columns with null values

X134 Columns with null values

X135 Columns with null values

X136 Columns with null values

X137 Columns with null values

X138 Columns with null values

X139 Columns with null values

X140 Columns with null values

X141 Columns with null values

X142 Columns with null values

X143 Columns with null values

X144 Columns with null values

X145 Columns with null values

X146 Columns with null values

X147 Columns with null values

X148 Columns with null values

X149 Columns with null values

X150 Columns with null values

X151 Columns with null values

X152 Columns with null values

Source

Collected and entered by BSc (Hons) Statistics undegraduates - 2020

Examples

```
data(DSraw)
head(DSraw)
summary(DSraw)
```

DStidy	*Data scientists, data analyst, and statistician job advertisements from 2020 to 2023*

Description

A dataset with 1172 rows and 109 variables

Usage

```
data(DStidy)
```

Details

- ID. row id
- Consultant. Name of the consultant
- DateRetrieved. Date of Data Retrieved
- DatePublished. Published Date of the Advertisement
- Job_title. Name of the job category
- Company. Name of the Company
- R. If R is required -> 1 ,If not mentioned -> 0
- SAS. If SAS is required -> 1 , If not mentioned -> 0
- SPSS. If SPSS is required -> 1 , If not mentioned -> 0
- Python. If Python is required -> 1 , If not mentioned -> 0
- MAtlab. If Matlab is required -> 1 , If not mentioned -> 0
- Scala. If Scala is required -> 1 , If not mentioned -> 0
- C#. If C# is required -> 1 , If not mentioned -> 0
- MS Word. If knowledge in MS Word is required -> 1 , If not mentioned -> 0
- Ms Excel. If knowledge in MS Excel is required -> 1 , If not mentioned -> 0
- OLE/DB. If knowledge in OLE/DB is required -> 1 , If not mentioned -> 0
- Ms Access. If Ms Access is required -> 1 , If not mentioned -> 0
- Ms PowerPoint. If knowledge in Ms Powerpoint is required -> 1 , If not mentioned -> 0
- Spreadsheets. If knowledge in Spreadsheets is required -> 1 , If not mentioned -> 0
- Data_visualization. If knowledge in Data Visualization is required -> 1 , If not mentioned -> 0
- Presentation_Skills. If Presentation Skills are required -> 1 , If not mentioned -> 0
- Communication. If Communication skills are required -> 1 , If not mentioned -> 0
- BigData. If knowledge in Big Data analysis is required -> 1 , If not mentioned -> 0
- Data_warehouse. If knowledge in Data Warehouse is required -> 1 , If not mentioned -> 0
- cloud_storage. If knowledge in Cloud Storage is required -> 1 , If not mentioned -> 0

- Google_Cloud. If knowledge in Google Cloud is required -> 1 , If not mentioned -> 0
- AWS. If knowledge in AWS is required -> 1 , If not mentioned -> 0
- Machine_Learning. If knowledge in Machine Learning is required -> 1 , If not mentioned -> 0
- Deep Learning. If knowledge in Deep Learning is required -> 1 , If not entioned -> 0
- Computer_vision. If knowledge in Computer Vision is required -> 1 , If not mentioned -> 0
- Java. If Java is required -> 1 , If not mentioned -> 0
- C++. If C++ is required -> 1 , If not mentioned -> 0
- C. If C is required -> 1 , If not mentioned -> 0
- Linux/Unix. If knowledge in Linux/Unix is required -> 1 , If not mentioned -> 0
- SQL. If SQL is required -> 1 , If not mentioned -> 0
- NoSQL. If NoSQL is required -> 1 , If not mentioned -> 0
- RDBMS. If knowledge in RDBMS is required -> 1 , If not mentioned -> 0
- Oracle. If knowledge in Oracle is required -> 1 , If not mentioned -> 0
- MySQL. If MYSQL is required -> 1 , If not mentioned -> 0
- PHP. If PHP is required -> 1 , If not mentioned -> 0
- Flash_Actionscript. If knowledge in Flash Action Script is required -> 1 , If not mentioned -> 0
- SPL. If knowledge in SPL is required -> 1 , If not mentioned -> 0
- web_design_and_development_tools. If knowledge in Web Design and Development Tools is required -> 1 , If not mentioned -> 0
- Wordpress. If knowledge in Wordpress is required -> 1 , If not mentioned -> 0
- AI. If Artificial Intelligence is required -> 1 , If not mentioned -> 0
- Natural_Language_Processing(NLP). If knowledge in NLP is required -> 1 , If not mentioned -> 0
- Microsoft Power BI. If knowledge in Microsoft Power BI is required -> 1 , If not mentioned -> 0
- Google_Analytics. If knowledge in Google Analytics is required -> 1 , If not mentioned -> 0
- graphics_and_design_skills. If Graphic and Design Skills are required -> 1 , If not mentioned -> 0
- Data_marketing. If Data Marketing abillity is required -> 1 , If not mentioned -> 0
- SEO. If knowledge in SEO is required -> 1 , If not mentioned -> 0
- Content_Management. If knowledge in Content Management is required -> 1 , If not mentioned -> 0
- Tableau. If knowledge in Tableau is required -> 1 , If not mentioned -> 0
- D3. If knowledge in D3 is required -> 1 , If not mentioned -> 0
- Alteryx. If knowledge in Alteryx is required -> 1 , If not mentioned -> 0
- KNIME. If knowledge in KNIME is required -> 1 , If not mentioned -> 0
- Spotfire. If knowledge in Spotfire is required -> 1 , If not mentioned -> 0

- Spark. If knowledge in Spark is required -> 1 , If not mentioned -> 0
- S3. If knowledge in S3 is required -> 1 , If not mentioned -> 0
- Redshift. If knowledge in Redshift is required -> 1 , If not mentioned -> 0
- DigitalOcean. If knowledge in Digital Ocean is required -> 1 , If not mentioned -> 0
- Javascript. If Java Script is required -> 1 , If not mentioned -> 0
- Kafka. If knowledge in Kafka is required -> 1 , If not mentioned -> 0
- Storm. If knowledge in Storm is required -> 1 , If not mentioned -> 0
- Bash. If knowledge in Bash is required -> 1 , If not mentioned -> 0
- Hadoop. If knowledge in Hadoop is required -> 1 , If not mentioned -> 0
- Data_Pipelines. If knowledge in Data Pipelines is required -> 1 , If not mentioned -> 0
- MPP_Platforms. If MPP Platforms is required ->1,If not mentioned-0
- Qlik. If Qlik is required ->1,If not mentioned ->0
- Pig. If Pig is required ->1,If not mentioned ->0
- Hive. If Hive is required ->1,If not mentioned ->0
- Tensorflow. If Tensorflow is required ->1,If not mentioned ->0
- Map/Reduce. If Map/Reduce is required ->1,If not mentioned ->0
- Impala. If Impala is required ->1,If not mentioned ->0
- Solr. If Sloris required ->1,If not mentioned ->0
- Teradata. If Teradata is required ->1,If not mentioned ->0
- MongoDB. If MonoDB is required ->1,If not mentioned ->0
- Elasticsearch. If Elasticsearch is required ->1,If not mentioned ->0
- YOLO. If YOLO is required-1 ,If not mentioned-0
- agile execution. If agile execution is required->1 ,If not mentioned->0
- Data_management. If the knowledge in data management is required->1 ,If not mentioned->0
- pyspark. If pyspark is required->1 ,If not mentioned->0
- Data_mining. If the knowledge in data mining is required->1 ,If not mentioned->0
- Data_science. If the knowledge in data science is required->1 ,If not mentioned->0
- Web_Analytic_tools. If the knowledge in Web Analytic tools is required->1 ,If not mentioned->0
- IOT. If IOT is required->1 ,If not mentioned->0
- Numerical_Analysis. If the knowledge in Numerical Analysis is required->1 ,If not mentioned->0
- Economic. If the knowledge in Economic is required->1 ,If not mentioned->0
- Finance_Knowledge. If Finance_Knowledge is required->1 ,If not mentioned->0
- Investment_Knowledge. If Investment Knowledge is required->1 ,If not mentioned->0
- Problem_Solving. If the ability of Problem Solving is required->1 ,If not mentioned->0
- Team_Handling. If the ability of Team Handling is required->1 ,If not mentioned->0

- Debtor_reconcilation. If the ability of Debtor reconcilation is required->1 ,If not mentioned->0

- Payroll_management. If Payroll management is required->1 ,If not mentioned->0

- Bayesian. If Bayesian is required->1 ,If not mentioned->0

- Optimization. If Optimization knowledge is required-1 ,If not mentioned-0

- Knowledge_in. Required knowledge to do a particular job ,If not mentioned->NA

- City. City where the company is located in

- Educational_qualifications. Required educational qualifications

- Salary. Amount of salary

- URL. Web address of a particular job advertisement

- Search_Term. web search term of a particular job advertisement

- Job_Category. Category of the job (i.e. "Data Science","Data Analyst" etc.)

- Team_Handling. If the ability of Team Handling is required-1 ,If not mentioned-0

- Debtor_reconcilation. If the ability of Debtor reconciliation is required-1 ,If not mentioned-0

- Payroll_management. If the ability of Payroll management is required-1 ,If not mentioned-0

- Bayesian. If Bayesian knowledge is required-1 ,If not mentioned-0

- Bahasa_Malaysia. If Bahasa Malaysia is required-1 ,If not mentioned-0

- English_proficiency. If English proficiency is required-1 ,If not mentioned-0

- Experience_Category. Number of years of experience in binned into categories

- Location. Location

- Payment Frequency. Payment frequency

- BSc_needed. If BSc is required-1 ,If not mentioned-0

- MSc_needed. If MSc is required-1 ,If not mentioned-0

- PhD_needed. If PhD is required-1 ,If not mentioned-0

- English Needed. If English is required-1 ,If not mentioned-0

- year. Survey year

Source

Data collection was done, BSc (Hons)Staistics, University of Sri Jayewardenepura under the statistical consultancy service from 2020 to 2023.

DStidy_2020	*Data scientists, data Analyst, and statistician related job advertisements in 2020*

Description

A dataset with 430 rows and 115 columns

Usage

```
data(DStidy_2020)
```

Details

- ID. Row id
- Consultant. Name of the consultant
- DateRetrieved. Date of data retrieved
- DatePublished. Published date of the advertisement
- Job_title. Name of the job category
- Company. Name of the company
- R. If R is required -> 1 , If not mentioned -> 0
- SAS. If SAS is required -> 1 , If not mentioned -> 0
- SPSS. If SPSS is required -> 1 , If not mentioned -> 0
- Python. If Python is required -> 1 , If not mentioned -> 0
- MAtlab. If MAtlab is required -> 1 , If not mentioned -> 0
- Scala. If Scala is required -> 1 , If not mentioned -> 0
- C_Sharp. If C_Sharp is required -> 1 , If not mentioned -> 0
- Ms_Excel. If Ms_Excel is required -> 1 , If not mentioned -> 0
- OLE_DB. If OLE_DB is required -> 1 , If not mentioned -> 0
- Ms_Access. If Ms_Access is required -> 1 , If not mentioned -> 0
- Ms_PowerPoint. If Ms_PowerPoint is required -> 1 , If not mentioned -> 0
- Spreadsheets. If Spreadsheets is required -> 1 , If not mentioned -> 0
- Data_visualization. If knowledge in Data Visualization is required -> 1 , If not mentioned -> 0
- Presentation_Skills. If Presentation Skills are required -> 1 , If not mentioned -> 0
- Communication. If Communication skills are required -> 1 , If not mentioned -> 0
- BigData. If knowledge in Big Data analysis is required -> 1 , If not mentioned -> 0
- Data_warehouse. If knowledge in Data Warehouse is required -> 1 , If not mentioned -> 0
- cloud_storage. If knowledge in Cloud Storage is required -> 1 , If not mentioned -> 0
- Google_Cloud. If knowledge in Google Cloud is required -> 1 , If not mentioned -> 0

- AWS. If knowledge in AWS is required -> 1 , If not mentioned -> 0
- Machine_Learning. If knowledge in Machine Learning is required -> 1 , If not mentioned -> 0
- Deep_Learning. If knowledge in Deep Learning is required -> 1 , If not mentioned -> 0
- Computer_vision. If knowledge in Computer Vision is required -> 1 , If not mentioned -> 0
- Java. If Java is required -> 1 , If not mentioned -> 0
- Cpp. If Cpp is required -> 1 , If not mentioned -> 0
- C. If C is required -> 1 , If not mentioned -> 0
- Linux_Unix. If knowledge in Linux/Unix is required -> 1 , If not mentioned -> 0
- SQL. If SQL is required -> 1 , If not mentioned -> 0
- NoSQL. If NoSQL is required -> 1 , If not mentioned -> 0
- RDBMS. If knowledge in RDBMS is required -> 1 , If not mentioned -> 0
- Oracle. If knowledge in Oracle is required -> 1 , If not mentioned -> 0
- MySQL. If MYSQL is required -> 1 , If not mentioned -> 0
- PHP. If PHP is required -> 1 , If not mentioned -> 0
- Flash_Actionscript. If Flash_Actionscript is required -> 1 , If not mentioned -> 0
- SPL. If knowledge in SPL is required -> 1 , If not mentioned -> 0
- web_design_and_development_tools. If knowledge in Web Design and Development Tools is required -> 1 , If not mentioned -> 0
- Wordpress. If Wordpress is required -> 1 , If not mentioned -> 0
- AI. If AI is required 1 , If not mentioned 0
- Natural_Language_Processing(NLP). If knowledge in NLP is required -> 1 , If not mentioned -> 0
- Microsoft_Power_BI. If knowledge in Microsoft Power BI is required -> 1 , If not mentioned -> 0
- Google_Analytics. If knowledge in Google Analytics is required -> 1 , If not mentioned -> 0
- graphics_and_design_skills. If Graphic and Design Skills are required -> 1 , If not mentioned -> 0
- Data_marketing. If Data Marketing abillity is required -> 1 , If not mentioned -> 0
- SEO. If knowledge in SEO is required -> 1 , If not mentioned -> 0
- Content_Management. If knowledge in Content Management is required -> 1 , If not mentioned -> 0
- Tableau. If knowledge in Tableau is required -> 1 , If not mentioned -> 0
- D3. If knowledge in D3 is required -> 1 , If not mentioned -> 0
- Alteryx. If knowledge in Alteryx is required -> 1 , If not mentioned -> 0
- KNIME. If knowledge in KNIME is required -> 1 , If not mentioned -> 0
- Spotfire. If knowledge in Spotfire is required -> 1 , If not mentioned -> 0
- Spark. If knowledge in Spark is required -> 1 , If not mentioned -> 0
- S3. If knowledge in S3 is required -> 1 , If not mentioned -> 0

- Redshift. If knowledge in Redshift is required -> 1 , If not mentioned -> 0
- DigitalOcean. If knowledge in Digital Ocean is required -> 1 , If not mentioned -> 0
- Javascript. If Java Script is required -> 1 , If not mentioned -> 0
- Kafka. If knowledge in Kafka is required -> 1 , If not mentioned -> 0
- Storm. If knowledge in Storm is required -> 1 , If not mentioned -> 0
- Bash. If knowledge in Bash is required -> 1 , If not mentioned -> 0
- Hadoop. If knowledge in Hadoop is required -> 1 , If not mentioned -> 0
- Data_Pipelines. If knowledge in Data Pipelines is required -> 1 , If not mentioned -> 0
- MPP_Platforms. If MPP Platforms is required -> 1 , If not mentioned -> 0
- Qlik. If Qlik is required -> 1 , If not mentioned -> 0
- Pig. If Pig is required -> 1 , If not mentioned -> 0
- Hive. If Hive is required -> 1 , If not mentioned -> 0
- Tensorflow. If Tensorflow is required -> 1 , If not mentioned -> 0
- Map_Reduce. If Map/Reduce is required -> 1 , If not mentioned -> 0
- Impala. If Impala is required -> 1 ,If not mentioned -> 0
- Solr. If Sloris required -> 1 , If not mentioned -> 0
- Teradata. If Teradata is required -> 1 , If not mentioned -> 0
- MongoDB. If MonoDB is required -> 1 , If not mentioned -> 0
- Elasticsearch. If Elasticsearch is required -> 1, If not mentioned -> 0
- YOLO. If YOLO is required -> 1, If not mentioned -> 0
- agile_execution. If agile execution is required -> 1 , If not mentioned -> 0
- Data_management. If the knowledge in Data Management is required -> 1 , If not mentioned -> 0
- pyspark. If pyspark is required -> 1 , If not mentioned -> 0
- Data_mining. If the knowledge in Data Mining is required -> 1 , If not mentioned -> 0
- Data_science. If the knowledge in Data Science is required -> 1 , If not mentioned -> 0
- Web_Analytic_tools. If the knowledge in Web Analytic tools is required -> 1 , If not mentioned -> 0
- IOT. If IOT is required -> 1 , If not mentioned -> 0
- Numerical_Analysis. If the knowledge in Numerical Analysis is required -> 1 , If not mentioned -> 0
- Economic. If the knowledge in Economic is required -> 1 , If not mentioned -> 0
- Finance_Knowledge. If Finance_Knowledge is required -> 1 , If not mentioned -> 0
- Investment_Knowledge. If Investment Knowledge is required -> 1 , If not mentioned -> 0
- Problem_Solving. If the ability of Problem Solving is required -> 1 , If not mentioned -> 0
- Korean_language. If the ability of Korean language is required -> 1 , If not mentioned -> 0
- Bash_Linux_Scripting. If Bash Linux Scripting is required -> 1 , If not mentioned -> 0
- Team_Handling. If the ability of Team Handling is required -> 1 , If not mentioned -> 0

- Debtor_reconcilation. If the ability of Debtor reconciliation is required -> 1 , If not mentioned -> 0
- Payroll_management. If the ability of Payroll management is required -> 1 , If not mentioned -> 0
- Bayesian. If Bayesian knowledge is required -> 1 , If not mentioned -> 0
- Optimization. If Optimization knowledge is required -> 1 ,If not mentioned -> 0
- Bahasa_Malaysia. If Bahasa_Malaysia knowledge is required -> 1 ,If not mentioned -> 0
- Knowledge_in. Required knowledge to do a particular job , If not mentioned -> NA
- City. City where the company is located in , If not mentioned -> NA
- Location. Country where the company is located in
- Educational_qualifications. Required educational qualifications
- Salary. Salary
- English_proficiency. English proficiency
- URL. URL of the job advertisement
- Search_Term. Search Term
- Job_Category. Name of the job category
- Minimum_Years_of_experience. Minimum years of experience needed for the job , If not mentioned -> NA
- Experience. Experience
- Experience_Category. Experience category
- Job_Country. Job country
- Edu_Category. Education category
- Minimum_Salary. Minimum salary
- Salary_BasisSalary. basis

Source

Data wrangling part was done by Janith C. Wanniarachchie, BSc (Hons)Staistics, University of Sri Jayewardenepura and description file was prepared by Randi Shashikala.

| DStidy_2021 | *Data Scientists/Data Analyst/ Statistician Job Advertisements in the year 2021* |

Description

Job advertisements collected in the year 2021

Usage

```
DStidy_2021
```

Format

A data frame with 382 rows and 115 columns

ID Row id

Consultant Name of the consultant

URL Web address of a particular job advertisement

Search_Term Web search term of a particular job advertisement

DateRetrieved Date of data retrieved

DatePublished Published date of the advertisement

Job_Field Name of the related job field

Job_title Name of the job category

Company Name of the company

Knowledge_in Required knowledge to do a particular job , If not mentioned -> NA

Minimum Experience in Years Minimum years of experience needed for the job , If not mentioned -> NA

City City where the company is located in , If not mentioned -> NA

Location Country where the company is located in

Educational_qualifications Required educational qualifications

Payment Frequency Payment basis of salary(i.e. "hourly","daily","monthly","yearly", "NA")

Currency Currency type of the salary

Salary Amount of salary

English Needed If English proficiency is required -> 1 , If not mentioned -> 0

English proficiency description Required level of English proficiency , If not mentioned -> NA

Additional_languages If other lanuages except English is required -> 1 , If not mentioned -> NA

AI If Artificial Intelligence is required -> 1 , If not mentioned -> 0

Natural_Language_Processing(NLP) If knowledge in NLP is required -> 1 , If not mentioned -> 0

Data_Pipelines If knowledge in Data Pipelines is required -> 1 , If not mentioned -> 0

Machine_Learning If knowledge in Machine Learning is required -> 1 , If not mentioned -> 0

Deep Learning If knowledge in Deep Learning is required -> 1 , If not mentioned -> 0

Computer_vision If knowledge in Computer Vision is required -> 1 , If not mentioned -> 0

Data_visualization If knowledge in Data Visualization is required -> 1 , If not mentioned -> 0

Data_warehouse If knowledge in Data Warehouse is required -> 1 , If not mentioned -> 0

BigData If knowledge in Big Data analysis is required -> 1 , If not mentioned -> 0

Data_management If the knowledge in Data Management is required -> 1 , If not mentioned -> 0

Data_mining If the knowledge in Data Mining is required -> 1 , If not mentioned -> 0

Data_science If the knowledge in Data Science is required -> 1 , If not mentioned -> 0

Bayesian If Bayesian knowledge is required -> 1 , If not mentioned -> 0

Optimization If Optimization knowledge is required -> 1 ,If not mentioned -> 0

Numerical_Analysis If the knowledge in Numerical Analysis is required -> 1 , If not mentioned -> 0

IOT If IOT is required -> 1 , If not mentioned -> 0

Data_translation If the knowledge in Data Translation is required -> 1 , If not mentioned -> 0

R If R is required -> 1 ,If not mentioned -> 0

SAS If SAS is required -> 1 , If not mentioned -> 0

SPSS If SPSS is required -> 1 , If not mentioned -> 0

Python If Python is required -> 1 , If not mentioned -> 0

MAtlab If Matlab is required -> 1 , If not mentioned -> 0

Scala If Scala is required -> 1 , If not mentioned -> 0

C# If C# is required -> 1 , If not mentioned -> 0

Java If Java is required -> 1 , If not mentioned -> 0

C++ If C++ is required -> 1 , If not mentioned -> 0

C If C is required -> 1 , If not mentioned -> 0

Bash If Bash is required -> 1 , If not mentioned -> 0

Tensorflow If Tensorflow is required -> 1 , If not mentioned -> 0

pyspark If pyspark is required -> 1 , If not mentioned -> 0

YOLO If YOLO is required -> , If not mentioned -> 0

MS Word If knowledge in MS Word is required -> 1 , If not mentioned -> 0

Ms Excel If knowledge in MS Excel is required -> 1 , If not mentioned -> 0

Ms Access If Ms Access is required -> 1 , If not mentioned -> 0

Ms PowerPoint If knowledge in Ms Powerpoint is required -> 1 , If not mentioned -> 0

Spreadsheets If knowledge in Spreadsheets is required -> 1 , If not mentioned -> 0

Google_Analytics If knowledge in Google Analytics is required -> 1 , If not mentioned -> 0

Microsoft Power BI If knowledge in Microsoft Power BI is required -> 1 , If not mentioned -> 0

Tableau If knowledge in Tableau is required -> 1 , If not mentioned -> 0

D3 If knowledge in D3 is required -> 1 , If not mentioned -> 0

Qlik If Qlik is required -> 1 , If not mentioned -> 0

KNIME If knowledge in KNIME is required -> 1 , If not mentioned -> 0

Spotfire If knowledge in Spotfire is required -> 1 , If not mentioned -> 0

Linux/Unix If knowledge in Linux/Unix is required -> 1 , If not mentioned -> 0

OLE/DB If knowledge in OLE/DB is required -> 1 , If not mentioned -> 0

SQL If SQL is required -> 1 , If not mentioned -> 0

NoSQL If NoSQL is required -> 1 , If not mentioned -> 0

RDBMS If knowledge in RDBMS is required -> 1 , If not mentioned -> 0

Oracle If knowledge in Oracle is required -> 1 , If not mentioned -> 0

MySQL If MYSQL is required -> 1 , If not mentioned -> 0

MongoDB If MonoDB is required -> 1 , If not mentioned -> 0

MPP_Platforms If MPP Platforms is required -> 1 , If not mentioned -> 0

SPL If knowledge in SPL is required -> 1 , If not mentioned -> 0

Alteryx If knowledge in Alteryx is required -> 1 , If not mentioned -> 0

Spark If knowledge in Spark is required -> 1 , If not mentioned -> 0

Kafka If knowledge in Kafka is required -> 1 , If not mentioned -> 0

Hadoop If knowledge in Hadoop is required -> 1 , If not mentioned -> 0

Pig If Pig is required -> 1 , If not mentioned -> 0

Hive If Hive is required -> 1 , If not mentioned -> 0

Map/Reduce If Map/Reduce is required -> 1 , If not mentioned -> 0

Impala If Impala is required -> 1 ,If not mentioned -> 0

Storm If knowledge in Storm is required -> 1 , If not mentioned -> 0

Google_Cloud If knowledge in Google Cloud is required -> 1 , If not mentioned -> 0

AWS If knowledge in AWS is required -> 1 , If not mentioned -> 0

cloud_storage If knowledge in Cloud Storage is required -> 1 , If not mentioned -> 0

S3 If knowledge in S3 is required -> 1 , If not mentioned -> 0

Redshift If knowledge in Redshift is required -> 1 , If not mentioned -> 0

DigitalOcean If knowledge in Digital Ocean is required -> 1 , If not mentioned -> 0

Teradata If Teradata is required -> 1 , If not mentioned -> 0

Solr If Sloris required -> 1 , If not mentioned -> 0

Elasticsearch If Elasticsearch is required -> 1 , If not mentioned -> 0

Presentation_Skills If Presentation Skills are required -> 1 , If not mentioned -> 0

Communication If Communication skills are required -> 1 , If not mentioned -> 0

Problem_Solving If the ability of Problem Solving is required -> 1 , If not mentioned -> 0

Team_Handling If the ability of Team Handling is required -> 1 , If not mentioned -> 0

agile execution If agile execution is required -> 1 , If not mentioned -> 0

Data_marketing If Data Marketing abillity is required -> 1 , If not mentioned -> 0

SEO If knowledge in SEO is required -> 1 , If not mentioned -> 0

graphics_and_design_skills If Graphic and Design Skills are required -> 1 , If not mentioned -> 0

Content_Management If knowledge in Content Management is required -> 1 , If not mentioned -> 0

Economic If the knowledge in Economic is required -> 1 , If not mentioned -> 0

Finance_Knowledge If Finance_Knowledge is required -> 1 , If not mentioned -> 0

Investment_Knowledge If Investment Knowledge is required -> 1 , If not mentioned -> 0

Debtor_reconciliation If the ability of Debtor reconciliation is required -> 1 , If not mentioned -> 0

Payroll_management If the ability of Payroll management is required -> 1 , If not mentioned -> 0

web_design_and_development_tools If knowledge in Web Design and Development Tools is required -> 1 , If not mentioned -> 0

PHP If PHP is required -> 1 , If not mentioned -> 0

Javascript If Java Script is required -> 1 , If not mentioned -> 0

Web_Analytic_tools If the knowledge in Web Analytic tools is required -> 1 , If not mentioned -> 0

BSc_needed If a BSc Degree is required -> Yes , If not mentioned -> No/NA

MSc_needed If a MSc Degree is required -> Yes , If not mentioned -> No/NA

PhD_needed If a Phd Degree is required -> Yes , If not mentioned -> No/NA

Country Country

country_code country code

Job_Category Job category

Source

Data wrangling part was done by Janith C. Wanniarachchie, BSc (Hons)Staistics, University of Sri Jayewardenepura and description file was prepared by Randi Shashikala.

---

get_data	*Get data from DSjobtracker for specific years or all the years combined into one dataset*

Description

The DSjobtracker dataset is updated each year through the Statistical Consultancy Service of University of Sri Jayewardenepura. In order to accommodate the structural changes of data this function provides the capability to get the dataset required either combined through out the years or data specific to each year.

Usage

```
get_data(year)
```

Arguments

year   can be either "all" or an year after 2020 (2020,2021,...,etc.) as a numeric value

Index