Package ‘DSjobtracker’

December 14, 2020

Type Package

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

Version 0.1.1

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

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

License CC BY 4.0

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

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

VignetteBuilder knitr

Imports utils

NeedsCompilation no

Author Thiyanga S. Talagala [aut, cre],
        Janith C. Wanniarachchi [aut],
        Chammika R De Mel [ctb] (Data Collection, Data Wrangling),
        Jayani Lakshika [ctb] (Data Collection, Data Wrangling Dashboard),
        Shashini De Silva [ctb] (Data Collection, Data Wrangling),
        Thimani Dananjana [ctb] (Data Collection, Data Wrangling),
        Hansani Piyumika [ctb] (Hexsticker design),
        R Wijesuriya [ctb] (Data Collection),
        M Fernando [ctb] (Data Collection),
        N. N Madushani [ctb] (Data Collection),
        S. N Madushani [ctb] (Data Collection),
        V. P. C De Mel [ctb] (Data Collection),
        O Ranathunga [ctb] (Data Collection)

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2020-12-14 10:00:02 UTC
R topics documented:

- DSraw .......................................................... 2
- DSstidy ...................................................... 7
- open_dashboard ............................................... 10

Index

---

| 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 ability 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
<table>
<thead>
<tr>
<th>Korean_language</th>
<th>If the ability of speaking Korean language is required-&gt;1 ,If not mentioned-&gt;0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bash/Linux Scripting</td>
<td>If Bash\ Linux Scripting is required-&gt;1 ,If not mentioned-&gt;0</td>
</tr>
<tr>
<td>Knowledge_in</td>
<td>Required knowledge to do a particular job ,If not mentioned-&gt;NA</td>
</tr>
<tr>
<td>Experience</td>
<td>Minimum experience required for a particular job</td>
</tr>
<tr>
<td>City</td>
<td>City where the company is located in</td>
</tr>
<tr>
<td>Location</td>
<td>Country where the company is located in</td>
</tr>
<tr>
<td>Educational qualifications</td>
<td>Required educational qualifications</td>
</tr>
<tr>
<td>Salary</td>
<td>Amount of salary</td>
</tr>
<tr>
<td>Team_Handling</td>
<td>If the ability of Team Handling is required-1 ,If not mentioned-0</td>
</tr>
<tr>
<td>Debtor_reconciliation</td>
<td>If the ability of Debtor reconciliation is required-1 ,If not mentioned-0</td>
</tr>
<tr>
<td>Payroll_management</td>
<td>If the ability of Payroll management is required-1 ,If not mentioned-0</td>
</tr>
<tr>
<td>Bayesian</td>
<td>If Bayesian knowledge is required-1 ,If not mentioned-0</td>
</tr>
<tr>
<td>Optimization</td>
<td>If Optimization knowledge is required-1 ,If not mentioned-0</td>
</tr>
<tr>
<td>Bahasa Malaysia</td>
<td>If Bahasa Malaysia is required-1 ,If not mentioned-0</td>
</tr>
<tr>
<td>English proficiency</td>
<td>If English proficiency is required-1 ,If not mentioned-0</td>
</tr>
<tr>
<td>URL</td>
<td>Web address of a particular job advertisement</td>
</tr>
<tr>
<td>Search_Term</td>
<td>web search term of a particular job advertisement</td>
</tr>
</tbody>
</table>

| 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 undergraduates - 2020

Examples

data(DSraw)
head(DSraw)
summary(DSraw)
## Description

Job advertisements

## Usage

DStidy

## Format

A data frame with 430 rows and 114 variables

<table>
<thead>
<tr>
<th>ID</th>
<th>row id</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consultant</td>
<td>Name of the consultant</td>
</tr>
<tr>
<td>DateRetrieved</td>
<td>Date of Data Retrieved</td>
</tr>
<tr>
<td>DatePublished</td>
<td>Published Date of the Advertisement</td>
</tr>
<tr>
<td>Job_title</td>
<td>Name of the job category</td>
</tr>
<tr>
<td>Company</td>
<td>Name of the Company</td>
</tr>
<tr>
<td>R</td>
<td>If R is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>SAS</td>
<td>If SAS is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>SPSS</td>
<td>If SPSS is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>Python</td>
<td>If Python is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>Matlab</td>
<td>If Matlab is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>Scala</td>
<td>If Scala is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>C_Sharp</td>
<td>If C# is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>MS_Word</td>
<td>If knowledge in MS Word is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>Ms_Excel</td>
<td>If knowledge in MS Excel is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>OLE_DB</td>
<td>If knowledge in OLE/DB is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>Ms_Access</td>
<td>If Ms Access is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>Ms_PowerPoint</td>
<td>If knowledge in Ms Powerpoint is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>Spreadsheets</td>
<td>If knowledge in Spreadsheets is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>Data_visualization</td>
<td>If knowledge in Data Visualization is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>Presentation_Skills</td>
<td>If Presentation Skills are required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>Communication</td>
<td>If Communication skills are required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>BigData</td>
<td>If knowledge in Big Data analysis is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>Data_warehouse</td>
<td>If knowledge in Data Warehouse is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
<tr>
<td>cloud_storage</td>
<td>If knowledge in Cloud Storage is required -&gt; 1 , If not mentioned -&gt; 0</td>
</tr>
</tbody>
</table>
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 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 ability 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
<table>
<thead>
<tr>
<th>Knowledge Area</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>DigitalOcean</td>
<td>1</td>
</tr>
<tr>
<td>Javascript</td>
<td>1</td>
</tr>
<tr>
<td>Kafka</td>
<td>0</td>
</tr>
<tr>
<td>Storm</td>
<td>0</td>
</tr>
<tr>
<td>Bash</td>
<td>0</td>
</tr>
<tr>
<td>Hadoop</td>
<td>0</td>
</tr>
<tr>
<td>Data_Pipelines</td>
<td>0</td>
</tr>
<tr>
<td>Qlik</td>
<td>0</td>
</tr>
<tr>
<td>Pig</td>
<td>0</td>
</tr>
<tr>
<td>Hive</td>
<td>0</td>
</tr>
<tr>
<td>Tensorflow</td>
<td>0</td>
</tr>
<tr>
<td>Map_Reduce</td>
<td>0</td>
</tr>
<tr>
<td>Impala</td>
<td>0</td>
</tr>
<tr>
<td>Solr</td>
<td>0</td>
</tr>
<tr>
<td>Teradata</td>
<td>0</td>
</tr>
<tr>
<td>MongoDB</td>
<td>0</td>
</tr>
<tr>
<td>Elasticsearch</td>
<td>0</td>
</tr>
<tr>
<td>YOLO</td>
<td>0</td>
</tr>
<tr>
<td>agile_execution</td>
<td>0</td>
</tr>
<tr>
<td>Data_management</td>
<td>0</td>
</tr>
<tr>
<td>pyspark</td>
<td>0</td>
</tr>
<tr>
<td>Data_mining</td>
<td>0</td>
</tr>
<tr>
<td>Data_science</td>
<td>0</td>
</tr>
<tr>
<td>Web_Analytic_tools</td>
<td>0</td>
</tr>
<tr>
<td>IOT</td>
<td>0</td>
</tr>
<tr>
<td>Numerical_Analysis</td>
<td>0</td>
</tr>
<tr>
<td>Economic</td>
<td>0</td>
</tr>
<tr>
<td>Finance_Knowledge</td>
<td>0</td>
</tr>
<tr>
<td>Investment_Knowledge</td>
<td>0</td>
</tr>
<tr>
<td>Problem_Solving</td>
<td>0</td>
</tr>
<tr>
<td>Korean_language</td>
<td>0</td>
</tr>
<tr>
<td>Bash_Linux_Scripting</td>
<td>0</td>
</tr>
<tr>
<td>Knowledge_in</td>
<td>NA</td>
</tr>
<tr>
<td>City</td>
<td>NA</td>
</tr>
<tr>
<td>Location</td>
<td>NA</td>
</tr>
</tbody>
</table>
**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
**Job_Category**  Category of the job (i.e. "Data Science","Data Analyst" etc.)
**Minimum_Years_of_experience**  Minimum years of experience needed for the job
**Experience**  Minimum experience required for a particular job
**Experience_Category**  Number of years of experience in binned into categories
**Job_Country**  Country where job is located in
**Edu_Category**  Type of educational qualification needed
**Minimum_Salary**  Starting salary proposed in respective currency
**Salary_Basis**  Payment basis of salary (i.e. "unspecified","hourly","daily","monthly","yearly")

**Source**

Data wrangling part was done by Chammika R. De Mel, Shashini D. Silva and Thimani D Ranathunga, BSc (Hons)Staistics, University of Sri Jayewardenepura.

**Examples**

```r
data(DStidy)
head(DStidy)
summary(DStidy)
```

---

**open_dashboard**  
**Visualize the dataset in a dashboard**

---

**Description**

Opens the dashboard containing visualizations of the dataset.

The dashboard is hosted at [https://scs-fas-sjp.netlify.app/dsjobtrackervisualisation/dataviz](https://scs-fas-sjp.netlify.app/dsjobtrackervisualisation/dataviz)
open_dashboard

Usage

open_dashboard()

Details

The data wrangling dashboard was created by Jayani Lakshika

Examples

open_dashboard()
Index

* datasets
  DSraw, 2
  DStidy, 7

DSraw, 2
DStidy, 7

open_dashboard, 10