

Data Science Tools and Ecosystem

In this notebook, Data Science Tools and Ecosystem are summarized.

Data Science Languages:

1. Python
2. R
3. SQL
4. Julia
5. Java
6. Ruby
7. Perl
8. Scala

Libraries commonly used in data science

1. Python Libraries

- Pandas
- Numpy
- SciPy
- Scikit-learn
- TensorFlow
- PyTorch
- Keras

2. R Libraries

- ggplot2
- dplyr
- tidyr
- caret
- randomForest
- shiny
- xgboost
- data.table

1. Julia Libraries:

- DataFrames.jl
- Flux.jl
- Gadfly.jl
- StatsBase.jl
- LightGBM.jl

Table of Data Science tools

Tool	language	Primary Function
Jupyter Notebook	Python	Interactive coding, data visualization, and documentation
Rstudio	R	Data analysis, statistical modeling, and visualization
Apache Spark	Scala, Python, R	Distributed data processing and analytics
Hadoop	Java, Python	Distributed data storage and processing
Excel	-	Spreadsheet-based data analysis and visualization
Google Colab	Python	Cloud-based Jupyter notebooks with GPU support

Below are a few examples of evaluating arithmetic expressions in Python

```
In [1]: # This a simple arithmetic expression to mutiPLY then add integers
(3*4)+5
```

Out[1]: 17

```
In [2]: # This will convert 200 minutes to hours by diving by 60.
minutes = 200
hours = minutes / 60
print(str(minutes) + " is equal to " + str(hours) + " hours")
```

200 is equal to 3.3333333333333335 hours

Objectives:

- List popular languages for Data Science
- Introduce key libraries for data analysis and machine learning
- Understand essential data science tools and their functions
- Learn to perform basic data manipulations and visualizations

Author

Sairam Pimple

