

Data Science Tools and Ecosystem

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

Objectives:

- List popular programming languages used by Data Scientists
- List popular code libraries for Data Science
- List Open source tools for Data Science
- Show examples of using Python to evaluate arithmetic expressions

Some of the popular languages that Data Scientists use are:

1. Python
2. R
3. SQL
4. C++
5. Java
6. Scala
7. Julia

Some of the commonly used libraries used by Data Scientists include:

1. NumPy: based on arrays and matrices, and is good for mathematical operations.
2. Pandas: data manipulation, cleaning, analysis. Built on top of NumPy
3. Matplotlib: graphing, most popular.
4. Seaborn: based on Matplotlib, for heat-maps, time series and violin charts

For Machine learning, the most popular options are:

1. Scikit Learn: for regression, clustering and classification models.
2. Keras: allows building the standard Deep Learning model.
3. TensorFlow: Deep Learning models production and deployment, experimentation.
4. PyTorch: for regression and classification models.

Data Science Tools

Jupyter Notebooks or JupyterLab

RStudio

Apache Zeppelin

Other non-open source options include Visual Studio Code (code is OSS but software is under the Microsoft Software License), Google Collab, and more.

Below are a few examples of evaluating arithmetic expressions in Python

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

```
Out[ ]: 17
```

```
In [ ]: # This will convert 200 minutes to hours by diving by 60  
200 / 60
```

```
Out[ ]: 3.3333333333333335
```

Author

Itamar Shiran

Thanks dude!! :D