Data Science Tools and Ecosystem

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

Objectives:

- List popular languages for Data Science.
- Introduce arithmetic expressions and their evaluation in Python.
- Demonstrate converting minutes to hours using Python.
- Explore basic data science tools and development environments.
- Provide examples of evaluating arithmetic expressions in Python.

Some of the popular languages that Data Scientists use are:

```
1. Python: Python is one of the most widely used languages in the field of data science. It
offers a rich ecosystem of libraries and frameworks such as NumPy, Pandas, Matplotlib, and
Scikit-learn, which provide powerful tools for data manipulation, analysis, visualization, and
machine learning.

2. R: R is another language commonly used by Data Scientists. It is particularly known for its
extensive collection of statistical and graphical techniques. R provides a wide range of
packages like ggplot2, dplyr, and caret, which enable efficient data manipulation,
visualization, and modeling.

3. Julia: Julia is an emerging programming language specifically designed for high-performance
numerical computing and scientific computing. It combines the ease of use of languages like
Python with the speed of languages like C or Fortran. Julia has gained popularity among data
scientists due to its ability to quickly perform complex calculations while maintaining
readable syntax.
```

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

```
1. NumPy: NumPy is a fundamental library for numerical computing in Python. It provides
support for large, multi-dimensional arrays and matrices, along with a collection of
mathematical functions to operate on these arrays efficiently. NumPy is widely used for tasks
such as data manipulation, linear algebra, statistical analysis, and more in data science.

2. Pandas: Pandas is a powerful data manipulation and analysis library. It provides easy-to-
use data structures, such as DataFrame, for handling structured data. Pandas allows you to
clean, transform, and analyze data effectively. It offers various functionalities like data
filtering, merging, grouping, and time series manipulation. Pandas is extensively used for
data preprocessing and exploratory data analysis (EDA).

3. Scikit-learn: Scikit-learn is a popular machine learning library in Python. It provides a
wide range of supervised and unsupervised learning algorithms, including classification,
regression, clustering, dimensionality reduction, and model selection. Scikit-learn also
offers utilities for data preprocessing, model evaluation, and hyperparameter tuning. It is
known for its user-friendly API and extensive documentation, making it suitable for both
beginners and experienced data scientists.

4. TensorFlow: TensorFlow is an open-source machine learning framework developed by Google. It
is widely used for building and training deep learning models. TensorFlow offers a flexible
architecture that allows you to define computational graphs and perform efficient numerical
computations. It supports various neural network architectures, including convolutional neural
networks (CNNs) and recurrent neural networks (RNNs). TensorFlow is favored by researchers and
practitioners in the field of deep learning.
```

Data Science Tools
Jupyter Notebook
RStudio
PyCharm Community Edition

Below are a few examples of evaluating arithmetic expressions in Python

```python
In [5]:   # This is a simple arithmetic expression to multiply then add integers.
          result = (3 * 4) + 5
          result
```

```
Out[5]:   17
```

```python
In [6]:   # This will convert 200 minutes to hours by dividing by 60.
```

```
minutes = 200
hours = minutes / 60
hours
```

Out[6]: 3.3333333333333335

Author

Khagendra Samal

In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js