

Python Basics for Data Science

Introduction to Python

Python is one of the most popular languages for data science due to its simplicity, readability, and extensive ecosystem of libraries. In this guide, we'll cover the essential Python basics that are frequently used in data science.

1. Python Basics

Variables and Data Types

Python supports different data types including integers, floats, strings, and booleans. You can declare variables without specifying their type, as Python infers the type.

Example:

```
x = 5 # Integer
```

```
y = 3.14 # Float
```

```
name = 'Data Science' # String
```

Lists, Tuples, and Dictionaries

Python has several data structures to store collections of data:

- List: A mutable, ordered sequence of items.
- Tuple: An immutable, ordered sequence of items.
- Dictionary: A collection of key-value pairs.

Example:

```
lst = [1, 2, 3]
```

```
tup = (1, 2, 3)
```

```
d = {'name': 'Python', 'version': 3.9}
```

Conditional Statements

Python uses if, elif, and else for conditionals.

Example:

if x > 0:

```
    print('Positive')
```

elif x == 0:

```
    print('Zero')
```

else:

```
    print('Negative')
```

2. Functions

Functions in Python are defined using the def keyword.

Example:

```
def add(a, b):
```

```
    return a + b
```

Lambda functions are anonymous functions defined in a single line.

Example:

```
add = lambda a, b: a + b
```

3. Python Libraries for Data Science

NumPy: NumPy is a library for numerical computing in Python. It provides support for large multi-dimensional arrays and matrices.

Example:

```
import numpy as np
```

```
arr = np.array([1, 2, 3])
```

Pandas: Pandas is a powerful data manipulation library. It provides two main data structures: Series and DataFrames.

Example:

```
import pandas as pd

df = pd.DataFrame({'A': [1, 2], 'B': [3, 4]})
```

Matplotlib/Seaborn: Libraries for data visualization.

Example:

```
import matplotlib.pyplot as plt

plt.plot([1, 2, 3], [4, 5, 6])
```

4. Data Manipulation with Pandas

Pandas allows easy manipulation of data. You can read data from files, inspect it, clean it, and perform various transformations.

Example:

```
df = pd.read_csv('data.csv')

print(df.head())
```

5. Data Visualization

Data visualization is essential for understanding patterns and trends in data.

Using Matplotlib and Seaborn, you can create various types of plots, including line charts, bar charts, histograms, and heatmaps.

Example:

```
sns.heatmap(df.corr(), annot=True)
```