08/10/2025

Libraries

A library in Python is a collection of modules and functions designed to perform specific tasks, helping you avoid rewriting common functionality.

Pandas- It is used for Data manipulation-used for handling and analyzing structured data (rows & columns)

Numpy- It is used for Numerical computing for multi-dimensional arrays and mathematical operations.

Statistics- It is used for Statistical operations for offers functions like mean, median, mode, stdev, etc.

Matplotlib, Seaborn, Plotly- It is used for Data visualization for creating various types of charts and plots.

Sklearn(Scikit-learn)-It is used for Machine Learning that contains algorithms for classification, regression, clustering, etc.

Keras- It is used for Deep Learning for High-level API for neural networks, usually runs on TensorFlow backend.

TensorFlow-It is used for Deep Learning framework it was developed by Google and it is used for neural networks and ML models.

PyTorch-It is used for Deep Learning framework it was developed by Facebook and it is flexible and widely used for research.

Importing and Installing Modules

import pandas as pd

- This imports the Pandas library and gives it an alias pd.
- You can now use pd to access Pandas functions like pd.DataFrame(), pd.Series(), etc.

Installing Modules

!pip install module_name

Avoiding Warnings

```
import warnings

warnings.filterwarnings('ignore')
```

Pandas

Pandas is a Python library used for data manipulation and analysis.

It provides two primary data structures:

1. **Series** → 1D labeled array (like a single column of data).
2. **DataFrame** → 2D labeled data structure (like an Excel sheet).

Series

```
s = pd.Series([10, 20, 30, 40, 50])

print(s)
```

- pd.Series() creates a Pandas Series.
- If no index is provided, it automatically assigns default integer indexes (0, 1, 2, …).

Reading Excel File

```
d1 = pd.read_excel(r"C:\Users\ emp_data.xlsx")
```

Reads an Excel file from the given path and stores it in a dataframe named d1.

Reading CSV File

```
d2 = pd.read_csv(r"C:\Users \mtcars.csv")
```

Reads a CSV file and stores it in dataframe d2

Reading Text File

```
d3 = pd.read_csv(r"C:\Users\ studata.txt")
```

Reads a text file and stores the content into dataframe d3

Reading TSV File

```
d4 = pd.read_csv(r"C:\Users\chipotle.tsv", sep="\t")
```

Reads a TSV (tab-separated values) file using sep="\t" and stores it in dataframe d4

Checking Dataset Shape

d2.shape

Displays the number of rows and columns present in the dataframe d2

Checking Number of Rows

d2.shape[0]

Shows only the number of rows in the dataframe.

Checking Number of Columns

d2.shape[1]

Shows only the number of columns in the dataframe.

View Top Records

d2.head()

d2.head(10)

Displays the first 5 rows by default or first 10 rows if the number is specified.

View Bottom Records

d2.tail()

d2.tail(10)

Displays the last 5 rows by default or last 10 rows if the number is specified.

View Random Records

d2.sample()

d2.sample(5)

Displays random records from the dataframe. By default, shows 1 row; with number, shows given count.