

FDS Lab assignment - 1

Name : Ayush Panchal

Roll No : P24DS013

Versions of modules installed

```
In [ ]: !python --version
```

Python 3.12.2

```
In [ ]: import nltk as nlt
import numpy as np
import scipy as sp
import pandas as pd
import matplotlib as mlt
import seaborn as sns
import sklearn
import plotly
```

```
In [ ]: print(f"NLTK Version : {nlt.__version__}")
print(f"Scipy Version : {sp.__version__}")
print(f"Pandas Version : {pd.__version__}")
print(f"Numpy Version : {np.__version__}")
print(f"Matplotlib Version : {mlt.__version__}")
print(f"Seaborn Version : {sns.__version__}")
print(f"Scikit-learn Version : {sklearn.__version__}")
print(f"Plotly Version : {plotly.__version__}")
```

NLTK Version : 3.9.1
Scipy Version : 1.14.1
Pandas Version : 2.2.2
Numpy Version : 1.26.4
Matplotlib Version : 3.9.2
Seaborn Version : 0.13.2
Scikit-learn Version : 1.5.1
Plotly Version : 5.23.0

NLTK Utilities and usage

Utility:

Provides a collection of text processing tools and corpora for tasks like tokenization, stemming, tagging, parsing, and semantic reasoning.

Use Cases:

- Text classification (e.g., spam filtering, sentiment analysis)

- Information extraction (e.g., named entity recognition, relationship extraction)
- Machine translation
- Question answering
- Text summarization

NumPy (Numerical Python)

Utility:

Fundamental package for scientific computing, providing multi-dimensional arrays and matrices, along with mathematical functions to operate on them efficiently.

Use Cases:

- Linear algebra operations
- Fourier transforms
- Random number generation
- Data manipulation and analysis
- Scientific simulations

Pandas

Utility:

Provides high-performance data structures (Series and DataFrame) and data analysis tools for handling and manipulating structured data.

Use Cases:

- Data cleaning and preprocessing
- Data exploration and analysis
- Data visualization
- Data manipulation and transformation

SciPy

Utility:

Built on NumPy, SciPy offers a collection of algorithms for optimization, integration, interpolation, linear algebra, and more.

Use Cases:

- Scientific and technical computing
- Signal processing
- Image processing
- Optimization and root-finding
- Statistical analysis

Matplotlib

Utility:

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. It provides a range of tools to generate various types of plots and charts.

Use Cases:

- Data visualization and exploration
- Creating static, animated, and interactive plots
- Generating line plots, scatter plots, bar charts, histograms, and more
- Customizing plot appearance with colors, labels, and legends
- Creating publication-quality graphics and visualizations

Scikit-learn (sklearn)

Utility:

Scikit-learn is a powerful Python library for machine learning and data mining. It provides simple and efficient tools for data analysis and modeling, including classification, regression, clustering, and dimensionality reduction algorithms.

Use Cases:

- Machine learning model development and evaluation
- Supervised learning (e.g., classification and regression)
- Unsupervised learning (e.g., clustering and dimensionality reduction)
- Feature selection and extraction
- Model evaluation and validation techniques (e.g., cross-validation)
- Pipelines for automated workflows and preprocessing

Plotly

Utility:

Plotly is an interactive graphing library that makes creating data visualizations easy, offering support for a wide variety of chart types. It integrates well with Python, R, and other programming languages and provides rich visualizations with interactivity.

Use Cases:

- Creating interactive dashboards and web applications
- Plotting data with complex 3D and 2D visualizations
- Visualizing geospatial data (e.g., choropleth maps)
- Performing real-time data analysis with streaming data visualizations
- Enhancing business reports with interactive charts

Seaborn

Utility:

Seaborn is a Python data visualization library built on top of Matplotlib, designed for making attractive and informative statistical graphics. It simplifies the process of creating complex visualizations like categorical plots, distribution plots, and regression plots.

Use Cases:

- Exploring and visualizing data distributions
- Creating pair plots and multi-plot grids to visualize relationships between multiple variables
- Enhancing statistical analysis with visual aids such as correlation heatmaps
- Plotting time series data for trend analysis
- Styling plots with built-in themes for consistency and clarity