

# Python Libraries for Data Science

## 1. Core Foundations

- NumPy – Numerical computing and array manipulation.
- Pandas – Data cleaning, analysis, and manipulation.
- Matplotlib – Basic data visualization.
- Seaborn – Statistical and aesthetic visualization.

## 2. Data Processing & Manipulation

- Scipy – Advanced math, stats, and optimization.
- Dask – Parallel & large-scale data processing.
- PyArrow – Fast data interchange (Parquet files).
- OpenPyXL / xlrd – Excel data manipulation.

## 3. Machine Learning & Predictive Modeling

- Scikit-learn – Traditional ML algorithms.
- XGBoost – Gradient boosting.
- LightGBM – Fast gradient boosting by Microsoft.
- CatBoost – Handles categorical data efficiently.
- Statsmodels – Statistical analysis and inference.

## 4. Data Visualization & Dashboards

- Plotly – Interactive plots and dashboards.
- Bokeh – Web-based interactive charts.
- Altair – Declarative plotting syntax.
- Dash / Streamlit – Building and deploying data apps.

## 5. Data Cleaning & Feature Engineering

- BeautifulSoup / lxml – Web scraping.
- Regex (re) – Text pattern matching.
- TextBlob / NLTK / SpaCy – NLP preprocessing.
- FeatureTools – Automated feature engineering.

## 6. Deep Learning & AI

- TensorFlow – Deep learning framework.
- Keras – High-level deep learning API.
- PyTorch – Deep learning for research and production.
- FastAI – Simplified deep learning on PyTorch.
- Hugging Face Transformers – Pretrained NLP models.

## **7. Time Series & Forecasting**

- Prophet – Forecasting by Facebook.
- statsmodels.tsa – Classical time-series models (ARIMA, SARIMA).
- sktime – Unified ML for time series.

## **8. Big Data, Cloud & Databases**

- PySpark – Big data processing.
- HDFS / Hadoop APIs – Data lake management.
- SQLAlchemy – Database ORM for Python.
- Boto3 / Google Cloud SDK – Cloud ML pipelines.

## **9. Automation & Deployment**

- MLflow – Model tracking and deployment.
- ONNX – Model format interchange.
- Joblib / Pickle – Save and load ML models.
- Docker SDK / FastAPI / Flask – API deployment.

## **10. Reporting & Documentation**

- Jupyter Notebook / JupyterLab – Interactive analysis.
- nbconvert / nbformat – Export notebooks to PDF/HTML.
- Papermill – Automate data pipelines.