Course Title: Data Science with Python basics

Course Duration: 50 hours

Course Objective:

This course aims to provide students with the necessary skills and tools to analyze real-world data, build models, and generate insights using Python. The course covers Python fundamentals, data manipulation, statistical analysis, machine learning, data visualization, and practical data science project work.

Week 1: Introduction to Data Science & Python

- **Topics:**
    - Overview of Data Science
    - The Data Science Process
    - Python Fundamentals for Data Science (variables, data types, loops, functions)
    - Introduction to Jupyter Notebooks
- **Assignments:**
    - Basic Python exercises (e.g., loops, functions, lists)

Week 2: Python Libraries for Data Science

- **Topics:**
    - Introduction to Python Libraries for Data Science
        - NumPy for numerical computing
        - Pandas for data manipulation
    - Basic Data Structures (Arrays, DataFrames, Series)
- **Assignments:**
    - NumPy and Pandas exercises (e.g., matrix operations, basic data manipulation)

Week 3: Data Wrangling and Cleaning

- **Topics:**
    - Importing and Exporting Data (CSV, Excel, SQL)
    - Handling Missing Data
    - Data Transformation and Cleaning
    - Exploratory Data Analysis (EDA)

- **Assignments:**
    - Cleaning and transforming real-world datasets

Week 4: Data Visualization

- **Topics:**
    - Matplotlib and Seaborn for Visualization
    - Creating Bar Charts, Line Plots, Scatter Plots, and Histograms
    - Advanced Plots: Heatmaps, Pairplots, Boxplots
    - Customizing Plots (color schemes, annotations)
- **Assignments:**
    - Visualization exercises using real datasets

Week 5: Probability and Statistics for Data Science

- **Topics:**
    - Descriptive Statistics (Mean, Median, Mode, Variance, Standard Deviation)
    - Probability Distributions (Normal, Binomial)
    - Hypothesis Testing (Z-test, T-test)
    - Introduction to Statistical Inference
- **Assignments:**
    - Statistical analysis of datasets

Week 6: Introduction to Machine Learning

- **Topics:**
    - Supervised vs. Unsupervised Learning
    - Introduction to Scikit-Learn
    - Simple Linear Regression
    - Model Evaluation (Train/Test Split, RMSE, MAE)
- **Assignments:**
    - Implementing a linear regression model on a dataset

Week 7: Classification Algorithms

- **Topics:**
    - Logistic Regression

    - Decision Trees and Random Forests
    - Evaluating Classification Models (Accuracy, Precision, Recall, F1 Score)
- **Assignments:**
    - Classification tasks using real datasets

Week 8: Clustering and Unsupervised Learning

- **Topics:**
    - Introduction to Clustering
    - K-Means Clustering
    - Dimensionality Reduction (PCA)
    - Hierarchical Clustering
- **Assignments:**
    - Applying clustering algorithms to datasets

Week 9: Advanced Machine Learning Techniques

- **Topics:**
    - Support Vector Machines (SVM)
    - Gradient Boosting (XGBoost, AdaBoost)
    - Hyperparameter Tuning (GridSearch, RandomizedSearch)
- **Assignments:**
    - Advanced machine learning models on real datasets

Week 10: Time Series Analysis

- **Topics:**
    - Introduction to Time Series Data
    - ARIMA Models
    - Seasonality and Trend Analysis
    - Forecasting
- **Assignments:**
    - Time series forecasting on historical data

Week 11: Deep Learning (Optional)

- **Topics:**

    - Introduction to Neural Networks
    - Basics of TensorFlow/Keras
    - Building a Simple Neural Network
    - Evaluating Neural Networks
- **Assignments:**
    - Building a basic neural network model

Week 12: Capstone Project

- **Topics:**
    - End-to-End Data Science Project
    - Problem Framing, Data Collection, and Cleaning
    - Model Building, Evaluation, and Reporting
    - Presentation of Results
- **Assignments:**
    - Complete a full data science project (including documentation and presentation)

Grading:

- Weekly Assignments: 40%
- Midterm Project: 20%
- Capstone Project: 30%
- Participation & Quizzes: 10%

Recommended Texts and Resources:

- **Books:**
    - *Python for Data Analysis* by Wes McKinney
    - *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* by Aurélien Géron
- **Online Resources:**
    - Kaggle Datasets and Competitions
    - Scikit-Learn Documentation
    - Pandas and NumPy Documentation