KenStack Data Science Course Syllabus (Phase-wise)

This Data Science course is designed for students and professionals to master the complete data science lifecycle — from data collection, preparation, analysis, and visualization to predictive modeling and deployment. The course includes hands-on projects and real-time case studies.

Phase 1: Foundations of Programming & Analytics

1. Python Programming for Data Science

Objectives:

- Learn the basics of Python programming required for data analysis and machine learning.

Topics Covered:

- Variables, Data Types, Operators
- Lists, Tuples, Dictionaries, Sets
- Conditional Statements, Loops
- Functions, Modules, File Handling
- Exception Handling, Comprehensions
- Introduction to OOPs

---

2. Statistics & Probability for Data Science

Objectives:

- Build a strong statistical foundation for data analysis.

Topics Covered:

- Types of Data & Sampling
- Measures of Central Tendency (Mean, Median, Mode)
- Measures of Dispersion (Range, Variance, Standard Deviation)
- Probability Concepts, Bayes Theorem
- Probability Distributions (Binomial, Normal, Poisson)
- Hypothesis Testing, t-test, z-test, Chi-square test

---

3. Excel for Data Analysis

Objectives:

- Analyze and visualize data using Microsoft Excel.

Topics Covered:

- Functions: VLOOKUP, IF, COUNTIF, etc.
- Pivot Tables, Charts, Data Cleaning
- Conditional Formatting, Dashboards

Phase 2: Data Analysis & Visualization

4. Data Analysis with Python

Objectives:

- Perform data cleaning, manipulation, and analysis using popular Python libraries.

Topics Covered:

- Numpy for Numerical Computation
- Pandas for Data Manipulation
- Handling Missing Values, Duplicates
- Data Aggregation & Grouping
- Merging & Joining DataFrames

---

5. Data Visualization with Python

Objectives:

- Create insightful and interactive visualizations.

Topics Covered:

- Matplotlib: Bar, Line, Pie, Histogram
- Seaborn: Box Plot, Heatmap, Pairplot
- Plotly (Optional): Interactive Visualizations
- Customizing Plots & Graphs

---

6. SQL for Data Science

Objectives:

- Learn to query and manipulate structured data using SQL.

Topics Covered:

- SELECT, WHERE, GROUP BY, HAVING, ORDER BY
- JOINS (INNER, LEFT, RIGHT, FULL)
- Subqueries, Aggregate Functions
- Creating & Modifying Tables
- Real-time SQL Practice on Sample Datasets

Phase 3: Machine Learning & Predictive Modeling

7. Introduction to Machine Learning

Objectives:

- Understand the concepts and workflows of machine learning.

Topics Covered:

- Supervised vs Unsupervised Learning
- ML Pipeline & Model Lifecycle
- Train/Test Split, Cross Validation

---

8. Supervised Learning Algorithms

Topics Covered:

- Linear Regression, Logistic Regression
- Decision Tree, Random Forest
- K-Nearest Neighbors (KNN)
- Support Vector Machine (SVM)

---

9. Unsupervised Learning Algorithms

Topics Covered:

- K-Means Clustering

- Hierarchical Clustering
- Dimensionality Reduction (PCA)

---

10. Model Evaluation & Tuning

Topics Covered:

- Confusion Matrix, Accuracy, Precision, Recall
- ROC & AUC Curve
- Hyperparameter Tuning (Grid Search, Random Search)

Phase 4: Project & Deployment

11. Real-Time Projects

Projects May Include:

- Sales Prediction Model
- Customer Segmentation
- Employee Attrition Prediction
- E-commerce Analytics Dashboard

---

12. Deployment Basics (Optional)

Topics Covered:

- Introduction to Streamlit or Flask
- Hosting a Model on Web
- Using GitHub for Version Control