

Data Science with R*

Duration - 20 Hours

Pricing – 20,000 INR

Module 1: Foundations

Objective: Set up R and gain a solid foundation in basic programming concepts.

Topics Covered:

- Install **R** and **RStudio**, and familiarize yourself with the IDE.
- Understand **variables**, **vectors**, **factors**, and **sequences**.
- Learn basic **loops**, **apply functions**, and foundational R programming techniques.

Module 2: Data Manipulation with R

Objective: Master data manipulation using key R packages.

Topics Covered:

- **dplyr:** Learn essential data manipulation functions like **select**, **filter**, **mutate**, **summarise**, and **group_by**.
- **tidyr:** Explore techniques for **gathering** and **spreading** data, and handling **missing data** effectively.

Module 3: Data Visualization

Objective: Create insightful visualizations using R tools.

Topics Covered:

- **ggplot2:** Understand **aesthetics**, **geoms**, **themes**, and **scales** to build custom visualizations.
- **plotly:** Learn how to create **interactive visualizations** for web applications and presentations.

Module 4: Statistical Analysis

Objective: Understand and perform basic statistical analysis.

Topics Covered:

- **Descriptive Statistics:** Summarize and understand data distributions.
- **Hypothesis Testing:** Conduct basic tests to validate assumptions.
- **Regression Analysis:** Understand simple and multiple linear regression techniques.

Module 5: Machine Learning with R

Objective: Learn to build and evaluate machine learning models.

Topics Covered:

- **caret:** Master model building, data preprocessing, and hyperparameter tuning.

- **Algorithms:** Implement **KNN**, **Random Forest**, **SVM**, and **K-Means Clustering**.
- **Model Validation:** Learn techniques like **cross-validation** and **confusion matrix** to evaluate model performance.

Module 6: Specialized Topics

Objective: Explore advanced topics in data science with R.

Topics Covered:

- **Time Series Analysis:** Use **ARIMA** and the **forecast** package for trend prediction.
- **Text Mining:** Learn how to process text data with **tm**, **tidytext**, **word clouds**, and **sentiment analysis**.
- **Big Data & AutoML:** Leverage **Spark in R** for handling big data and use **H2O** for deep learning models.

Module 7: SQL for Data Science in R

Objective: Integrate SQL with R for data querying and manipulation.

Topics Covered:

- Use R to interface with **SQL databases**.
- **SQL-based feature engineering** for machine learning models within R.

Module 8: Practical Projects

Objective: Apply your knowledge to real-world projects.

Projects Covered:

- **R Markdown Reports:** Automate data analysis and create reproducible reports.
- **Automated Workflows:** Use **R scripts** to streamline repetitive tasks.
- **Shiny App:** Develop an interactive **Shiny app** for customer segmentation and data visualization.