Lakshya institute of Technology

Data analysis

Studying data analytics is important because organizations rely on data to make smarter decisions. It offers strong career opportunities, high demand across industries, and global job options. Data analytics builds future-ready skills like problem-solving, critical thinking, and data interpretation, making it a valuable and sustainable career choice.

Module 1: Data Analysis using MS Excel

- Excel Environment & Workbook Basics, Data Types, Cell Formatting, Basic Excel Operations, Formulas & Cell References, Essential Functions: Math & Count, Logical, Text, Date, Sorting & Filtering, Data Validation, Tables & Structured Data, Basic Charts, Error Handling, Printing & Page Layout, File Import & Export
- Advance Excel: Advanced Functions, Logical & Nested Formulas, Math & Statistical Text manipulation functions, Excel tables and structured data, Pivot tables for data analysis, Advanced pivot table techniques, Advance charting techniques, Dashboard design principles, Advance dashboarding techniques

Module 2: Excel VBA

- Introduction to VBA & Excel Automation, VBA Programming Fundamentals, Control Structures, Working with Excel Objects, Data Processing & Automation, Functions & User Defined Functions (UDF), Events & Debugging, UserForms & Interaction, Advanced Automation, Optimization, Security & Best Practices

Module 3: Google Colab for data science

- Introduction to Google Colab, UI walkthrough.

Module 4: Exploratory Data Aanlysis(EDA)

- Introduction to EDA, Data Understanding & Inspection, Data Cleaning & Preprocessing, Univariate Analysis, Bivariate Analysis, Multivariate Analysis, Outlier Detection & Treatment, Feature Engineering & Transformation, Hypothesis Testing in EDA, EDA Reporting & Storytelling.

Module 5: Power BI basics

- Introduction to Power BI, Power BI Desktop Interface, Data Import & Connections, Power Query Basics (Data wrangling), Data Modeling Basics, DAX Fundamentals, Building Visuals, Basic Dashboard Design, Publishing & Sharing.

Module 6: Tableau

- Introduction to Tableau, Tableau Interface & Navigation, Data Connections & Preparation, Basic Visualizations, Calculations & Fields, Filters & Parameters, Advanced Visualizations, Dashboards & Storytelling, Performance & Optimization, Publishing & Collaboration.

Module 7: Pyspark (Apache Spark)

- Introduction, What is Apache Spark?, Apache Spark Architecture, Lazy Evaluation in Apache Spark, Spark Jobs, Stages, and Tasks, Databricks Free Account, Databricks Overview, Data Ingestion,

Databricks Notebook Overview, Spark Cluster, Data Reading with PySpark, Spark Data Reader API, Spark DAG, StructType and DDL Schema, Data Transformation with PySpark (Beginner Level), PySpark Intermediate Level Transformations, PySpark Advanced Level Functions, Window Functions in PySpark, User Defined Functions (UDF) in PySpark, Data Writing with PySpark, Data Writing Modes in PySpark, Parquet File Format, Managed vs External Tables in Spark, Spark SQL

Module 8: Apache Airflow

- What Apache Airflow is and when to use it (and when NOT to), Core concepts: DAGs Tasks Operators Hooks Sensors XCom, Airflow architecture explained with diagrams (Scheduler Webserver Executor Workers Metadata DB), Executors: Local Celery Kubernetes — when to use each, Writing DAGs using the modern TaskFlow API, Operators deep dive (Python Bash File Cloud operators), Variables Connections Secrets & best practices, XCom internals and anti-patterns, Sensors trigger rules and dependency management, Scheduling Cron Timetables Catchup & Backfill, Task Groups and Dynamic Task Mapping, Error handling retries SLAs logging & monitoring