

Data Analyst

A comprehensive learning
roadmap

Core Foundations

- Mathematics & Statistics for Data Analysis
- Probability & Distributions
- Descriptive & Inferential Statistics
- Linear Algebra Basics
- Critical Thinking & Problem Solving

Programming for Data Analysis

- Python for Data Analysis
- NumPy & Pandas
- Data Structures & Algorithms
- SQL & Relational Databases
- R Programming (Optional)

Data Wrangling & Preprocessing

- Data Cleaning Techniques
- Handling Missing Values & Outliers
- Feature Engineering
- Data Transformation & Normalization
- Working with APIs & Web Scraping

Data Visualization

- Matplotlib & Seaborn
- Plotly & Interactive Charts
- Tableau / Power BI
- Dashboard Design Principles
- Storytelling with Data

Databases & Data Engineering Basics

- SQL Advanced Queries (Joins, CTEs, Window Functions)
- NoSQL Databases (MongoDB, Cassandra)
- Data Warehousing Concepts
- ETL Pipelines
- BigQuery / Snowflake

Statistical Analysis & Experimentation

- Hypothesis Testing

- A/B Testing
- Regression Analysis
- Correlation & Causation
- Time Series Analysis

Machine Learning for Analysts

- Supervised & Unsupervised Learning
- Scikit-learn
- Classification & Regression Models
- Clustering (K-Means, DBSCAN)
- Model Evaluation Metrics

Business Intelligence & Analytics

- KPIs & Business Metrics
- Cohort Analysis
- Funnel Analysis
- Customer Segmentation
- Revenue & Growth Analytics

Tools & Ecosystem

- Jupyter Notebooks
- Git & Version Control
- Excel / Google Sheets (Advanced)
- Airflow (Pipeline Orchestration)
- dbt (Data Build Tool)

Cloud & Big Data

- AWS / Azure / GCP for Analytics
- Apache Spark & PySpark
- Google BigQuery & Looker
- Data Lakes vs Data Warehouses
- Streaming Data (Kafka Basics)

AI & LLM Integration

- Using LLMs for Data Summarization
- Prompt Engineering for Analysis
- AI-assisted EDA (Exploratory Data Analysis)
- AutoML Tools
- ChatGPT / Copilot for Productivity

Communication & Soft Skills

- Translating Data Insights to Stakeholders

- Report Writing & Documentation
- Presentation & Slide Design
- Data Ethics & Privacy (GDPR, CCPA)
- Domain Knowledge (Finance, Marketing, Healthcare)

Real-world Projects

- Sales Performance Dashboard (Power BI / Tableau)
- Customer Churn Prediction & Analysis
- E-commerce Funnel & Revenue Analysis
- COVID-19 / Public Dataset EDA Project
- End-to-End Analytics Pipeline with Python + SQL