

Introduction to Data Analytics

Introduction to Data Analytics

Quick reference guide for beginners

1. What Is Data Analytics?

Term	Meaning
Data Analytics	The systematic process of examining raw data to draw meaningful insights, support decision-making, and solve problems.
Analytics vs. Statistics	Statistics focuses on inference from samples; analytics adds the context of business goals, data engineering, and visualization.
Scope	Descriptive ? Diagnostic ? Predictive ? Prescriptive (the “analytics continuum”).

2. The Analytics Process (CRISP-DM Framework)

1. Business Understanding

- Define objectives, success criteria, and constraints.
- Example: “Increase customer churn prediction accuracy by 10%.”

2. Data Understanding

- Collect, explore, and assess data quality.
- Tools: SQL queries, data profiling, summary statistics.

3. Data Preparation

- Clean (handle missing values, outliers).
- Transform (normalize, encode categorical variables).
- Integrate (join tables, merge sources).

4. Modeling

- Choose appropriate algorithms (regression, classification, clustering, etc.).
- Split data (train / validation / test).

Introduction to Data Analytics

- Tune hyperparameters.

5. Evaluation

- Metrics aligned with business goals (accuracy, RMSE, AUC, ROI).
- Perform error analysis, cross-validation.

6. Deployment & Monitoring

- Export model (API, batch job).
- Set up alerts for data drift, performance decay.

(Many organizations iterate—feedback loops are essential.)

3. Types of Data Analytics

Type	Goal	Typical Techniques	Example Use Case
Descriptive	Summarize what happened	Aggregations, dashboards, KPI tables	Monthly sales report
Diagnostic	Explain why it happened	Correlation analysis, root cause analysis, drill downs	Why did sales drop in region X?
Predictive	Forecast future events	Regression, time series, classification, machine learning	Predict next quarter churn
Prescriptive	Recommend actions	Optimization, simulation, reinforcement learning	Allocate marketing budget for max ROI

4. Core Concepts & Tools

Concept	Description	Common Tools
Data Sources	Structured (databases, CSV), semi-structured (JSON, XML), unstructured (text, images)	SQL, NoSQL, APIs
Data Cleaning	Handling missing values, duplicates, outliers	Pandas, dplyr, OpenRefine

Introduction to Data Analytics

Concept	Description	Common Tools
Exploratory Data Analysis (EDA)	Visual & statistical inspection to discover patterns	Matplotlib, Seaborn, Tableau, Power BI
Feature Engineering	Creating informative variables (e.g., lag features, interaction terms)	Python (scikit?learn), R (caret)
Modeling	Training algorithms to capture relationships	scikit?learn, TensorFlow, XGBoost, SAS
Evaluation Metrics	Quantify model performance	Accuracy, Precision/Recall, F1, ROC?AUC, MAE, MSE
Visualization	Communicate findings clearly	ggplot2, Plotly, Power BI, Looker
Big Data Platforms	Process large volumes / velocity	Hadoop, Spark, Snowflake, Databricks
Version Control & Reproducibility	Track code, data, experiments	Git, DVC, MLflow, Jupyter