The **Data Science Process** typically follows these steps:

1. Problem Definition

- Understand the business or research problem
- Define objectives and success metrics
- Identify stakeholders and their requirements

2. Data Collection

- Gather raw data from multiple sources (databases, APIs, web scraping, sensors, etc.)
- Ensure data relevance and quality

3. Data Cleaning & Preprocessing

- Handle missing values, duplicate data, and inconsistencies
- Convert data into a structured format
- Perform feature engineering (creating new features from existing data)

4. Exploratory Data Analysis (EDA)

- Use statistical techniques and visualization (histograms, box plots, scatter plots, etc.)
- Identify patterns, correlations, and outliers
- Generate hypotheses for further analysis

5. Data Modeling

- Select appropriate machine learning (ML) or statistical models
- Train models using training data
- Tune hyperparameters for better performance

6. Model Evaluation

- Assess model performance using metrics like accuracy, precision, recall, F1-score, RMSE, etc.
- Compare different models and choose the best one
- Avoid overfitting and underfitting

7. Deployment

- Integrate the model into a production system (API, web app, or mobile app)

- Automate model predictions and decision-making

8. Monitoring & Maintenance

- Continuously track model performance over time
- Retrain models with new data if needed
- Handle concept drift and data drift

9. Communication & Visualization

- Present findings using reports, dashboards, and visualizations
- Interpret results in a meaningful way for stakeholders