

What Is the Data Science Lifecycle?

The **Data Science Lifecycle** is a structured sequence of stages that guide how data scientists **collect, process, analyze, model, and communicate** data-driven insights.

The 6 Key Stages of the Data Science Lifecycle

Stage	Description	Example
1. Problem Definition	Understand the business question or problem you want to solve.	“Can we predict customer churn?”
2. Data Collection	Gather data from multiple sources (databases, APIs, web scraping, etc.).	Collect customer usage data and complaints.
3. Data Cleaning & Preparation	Remove errors, handle missing values, and format data for analysis.	Fill missing age values, remove duplicates.
4. Exploratory Data Analysis (EDA)	Explore data visually and statistically to discover patterns.	Find which features relate most to churn.
5. Modeling & Evaluation	Build machine learning models and evaluate their performance.	Use logistic regression or random forest to predict churn.
6. Deployment & Communication	Deliver the model or results to production and share insights.	Deploy the model in a web app; report findings to management.