

NAAC Grade 'A'



**R. C. PATEL
INSTITUTE OF TECHNOLOGY**

An Autonomous Institute

MANTHAN

**Department of Computer Science &
Engineering (Data Science)**

**ISSUE YEAR
2022-23**



**“Data mirrors reality;
insight gives it
meaning.”**

**John Tukey
Father of Data
Science**



Vision-Mission

Institute Vision



To become a leading Institute in Technical education fostering innovation, research, ethical values, and sustainable development for the betterment of society.

Institute Mission



To impart high quality Technical Education through :

- Innovative and Interactive learning process and high quality, internationally recognized instructional programs.
 - Fostering a scientific temper among students by the means of a liaison with the Academia, Industries and Government.
 - Preparing students from diverse backgrounds to have aptitude for research and spirit of Professionalism.
 - To contribute to the nation's sustainable development.
- 



Vision-Mission

Department Vision



To provide cutting-edge Computer Engineering education in Data Science while instilling socio-moral values.

Department Mission



M1: To deliver state-of-the-art, ICT-enabled teaching and learning to achieve excellence in Data Science education.

M2: To develop professionally competent Data Science Engineers, meeting evolving industrial and societal needs.

M3: To prepare employable professionals with ethical values and a commitment to professional and social responsibility.



Visionary Insights

The Director



Prof. Dr. Jayantrao Bhaurao Patil

At RCPIT, we blend academic rigour with real-world exposure to create future-ready professionals. Emphasizing research, interdisciplinary learning, and soft skills, our hands-on approach empowers students to become ethical leaders, innovators, and entrepreneurs prepared to succeed in a rapidly evolving global landscape.

Visionary Insights

The Head of Department

.....



Prof. Dr. Ujwala Patil

Established in 2020 at RCPIT, Shirpur, the Computer Science & Engineering (Data Science) program trains students in algorithmic thinking and statistical modeling. This interdisciplinary field blends computer science, mathematics, and business analytics. Driven by massive industry demand, it is widely recognized as a highly dynamic and lucrative global career path.

Editor's Message



Prof. A. D. Mairale
Editor (Faculty)

DEAR READERS,

Welcome to the latest edition of our Department Magazine. This magazine is a platform to share knowledge, ideas, and achievements in Computer Science & Engineering (Data Science). It brings together expert articles, and career guidance from the Training and Placement Cell. We hope these pages inspire curiosity, encourage innovation, and motivate every student to learn continuously. May this magazine strengthen the bond between academia and industry while preparing our students for successful and rewarding careers.

Editorial Board



Mr. Vishal Bhalkar



Ms. Harshita Patil



Mr. Tejas Chaudhari



Contents

Section-I: Technical Articles



Why Every Engineer Needs Data Thinking

Statistics Isn't Boring — It's the Engine of Discovery

Exploratory Data Analysis: From Raw CSV to Insight

Data Collection in the Wild: Web Scraping and APIs

Section-II: Career Horizon



Careers in Data Science: What Roles Actually Exist?



Technical Article

Why Every Engineer Needs Data Thinking

Data is everywhere. It flows through every machine, every system, and every process. Engineers who understand data have a powerful edge. Traditionally, engineers focused on building things. Bridges, circuits, and software were the core goals. But the world has changed fast. Today, data shapes every engineering decision. Think about a simple manufacturing line. Sensors record pressure, temperature, and speed. Each reading is a data point. Without data thinking, these readings sit idle. With it, they reveal patterns and predict failures.

What Is Data Thinking?

Data thinking means asking the right questions about numbers. It means not just collecting data, but making sense of it. It is a skill, not just a tool. Consider a bridge engineer. She installs vibration sensors. Thousands of readings pour in each hour. Without data skills, she sees only noise. With them, she spots micro-fractures forming early. That insight saves lives. Data thinking combines curiosity, logic, and statistics. It asks: What does this number mean? Why did it change? What will happen next?

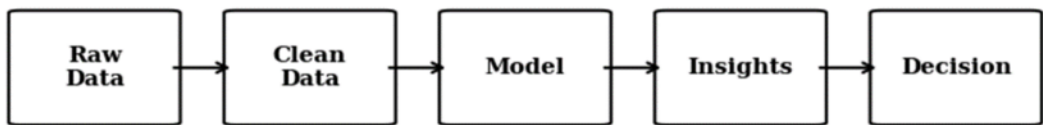


Fig. 1 — A simple data science workflow: from raw data to actionable decisions

Data Thinking in Real Engineering Jobs

Civil engineers now use data to plan road maintenance. They feed sensor readings into prediction models. The model predicts which road sections will fail soonest. Resources go where they are needed most. Electronics engineers use data to test circuit reliability. They collect failure logs over weeks. Statistical analysis reveals weak components. Products improve before they reach the market. Software engineers analyse user behaviour logs. They track click paths and load times. The data guides which features to build next. Product decisions become evidence-based, not guesswork.

Technical Article

Why Every Engineer Needs Data Thinking

The Skills Gap Is Real

A 2021 industry survey found something striking. Over 70% of Indian engineering firms said they lacked data-literate engineers. This gap costs companies time and money. It also creates a huge opportunity for new graduates.

Companies do not need every engineer to become a data scientist. They need engineers who can read charts, question assumptions, and work with data teams. These are learnable skills.

Starting Small

You do not need advanced mathematics to begin. Start with spreadsheets. Learn to filter and sort data. Ask what a chart is really showing.

Then move to Python. Libraries like Pandas and Matplotlib are friendly for beginners. They let you load, clean, and visualise data in a few lines of code.

Every engineering subject already contains data. Lab experiments generate measurements. Project reports contain figures. Practice treating those numbers as data, not just results.

A New Kind of Engineer

The engineer of the future is both a builder and a thinker. She designs systems and also analyses the data they produce. She uses evidence to improve her designs continuously.

Data thinking is not a replacement for engineering knowledge. It is an amplifier. It makes every other engineering skill more powerful and more impactful.

Start cultivating this mindset today. Your future workplace is already waiting for it.



Prof. Dr. Priti S. Sanjekar

Technical Article

Statistics Isn't Boring — It's the Engine of Discovery

Ask a room full of students what they think of statistics. Most will groan. It is seen as dry, difficult, and disconnected from real life. That perception is completely wrong. Statistics is the science of uncertainty. It tells us how confident we can be in any claim. Without it, data is just noise. With it, data becomes knowledge.

What Statistics Actually Does

When we collect data, we never get the full picture. We see a sample. Statistics helps us reason from that sample to the bigger truth behind it.

A doctor tests a new drug on 200 patients. She cannot test it on every patient in the world. Statistics lets her say, with measurable confidence, whether the drug works for everyone.

An engineer tests ten batches of steel. He wants to know if the whole production line is consistent. Statistics gives him the answer from just ten samples.

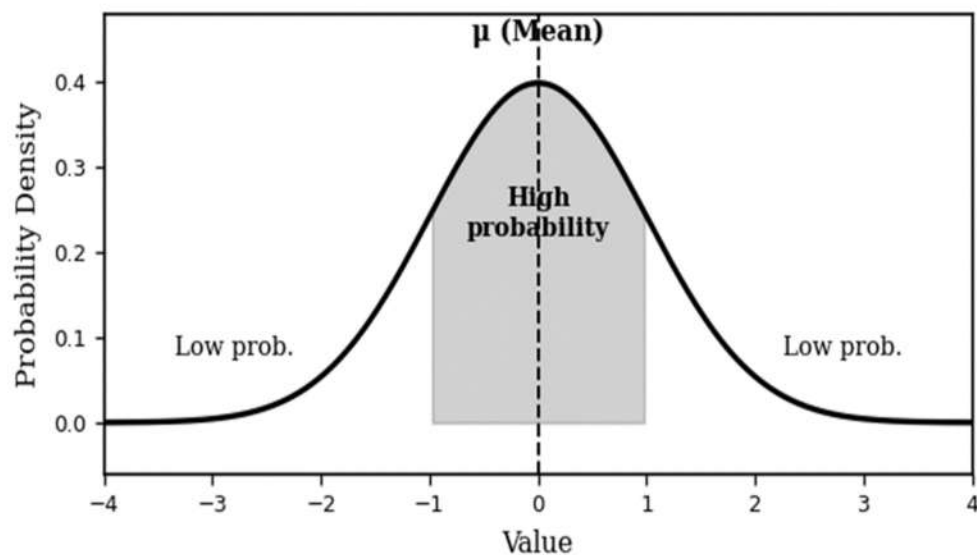


Fig. 2 — The normal distribution: most values cluster near the mean (μ)

Three Ideas That Change Everything

The first idea is the normal distribution. Many natural phenomena follow a bell-shaped curve. Heights, test scores, and measurement errors all cluster near a central average. Understanding this shape helps engineers set tolerances and predict outcomes.

Technical Article

Statistics Isn't Boring — It's the Engine of Discovery

The second idea is correlation vs causation. Just because two things move together does not mean one causes the other. Ice cream sales and swimming accidents both rise in summer. Neither causes the other. Confusing correlation with causation leads to terrible decisions. The third idea is hypothesis testing. We state a claim, collect data, and calculate the probability of seeing that data if the claim were false. If that probability is very low, we accept the claim as likely true. This is the foundation of scientific reasoning.

Statistics in Engineering Courses

You have already used statistics without realising it. When you calculated the mean of an experiment, that was statistics. When you drew an error bar on a graph, that was statistics too.

Quality control in manufacturing is built on statistical process control. Structural safety factors come from statistical failure analysis. Signal processing relies on probability distributions. Statistics is woven into every engineering domain.

Making It Intuitive

Statistics becomes clearer when you visualise it. Draw distributions by hand. Shade the area under the curve. Count how many data points fall within one standard deviation. That hands-on approach builds intuition faster than any formula.

Use real datasets. Government of India open data portals offer rainfall, crop yield, and census data for free. Load a dataset. Plot it. Ask what the shape tells you.

The Confidence to Question

Statistics gives you the courage to question claims. When you read that a material is "30% stronger," you will know to ask: stronger than what, and tested with what confidence level? That questioning mind is the mark of a good engineer and a good scientist.



Prof. A. D. Mairale

Technical Article

Exploratory Data Analysis: From Raw CSV to Insight

Every data project begins with a question. Before building any model, you must understand your data first. This process is called Exploratory Data Analysis, or EDA. It is the most important step in any data project.

EDA means looking at your data from every angle. You find patterns, spot anomalies, and check assumptions. Only then do you know what your data can actually tell you.

Step 1 — Load and Inspect

Open Python and import Pandas. Load your CSV file using `pd.read_csv()`. Immediately call `df.head()` to view the first five rows. Then run `df.info()`. This reveals column names, data types, and missing values at a glance.

Look for columns that are supposed to be numbers but are stored as text. That mismatch causes silent errors later. Fix data types early. It saves hours of debugging downstream.

Step 2 — Handle Missing Data

Real datasets are always messy. Sensors fail. Forms are left blank. Entries get corrupted. Missing values appear as `NaN` in Pandas. You must decide what to do with them.

The simplest option is to drop rows with missing values using `df.dropna()`. But if many rows have missing data, dropping them all can bias your analysis significantly.

A better option is imputation. Replace missing numbers with the column mean or median. For categorical values, use the most frequent entry. Pandas makes this straightforward using `fillna()`.

Step 3 — Visualise Everything

Numbers alone do not reveal patterns. Plots do. Use Matplotlib and Seaborn to visualise your data before doing anything else.

Histograms show how values are distributed. Are they clustered in one range? Are there outliers far from the main group? A quick histogram answers these questions in seconds.

Box plots show the median and spread. They highlight outliers as isolated dots. Correlation heatmaps show which columns move together. Strong correlations between input features and your target variable signal predictive power.

Technical Article

Exploratory Data Analysis: From Raw CSV to Insight

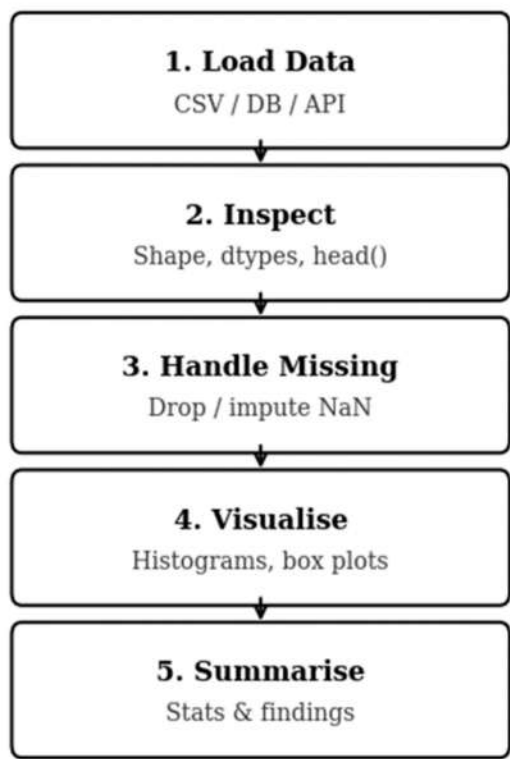


Fig. 3 — Five key steps in Exploratory Data Analysis



Prof. S. P. Salunkhe

Step 4 — Summarise with Statistics

Run `df.describe()` on your `DataFrame`. It returns count, mean, standard deviation, and quartiles for every numeric column. Scan this table carefully. A maximum salary value of ten million in a student dataset is probably a data entry error.

Groupby operations are equally powerful. Group data by a category and compute means. If average marks differ significantly between engineering branches, that group difference is a finding worth reporting.

EDA Is Not a Checklist

EDA is a conversation with your data. It is not a fixed sequence of steps. Sometimes a histogram raises a question that sends you back to the raw data. That is normal and healthy.

Experienced data scientists spend more time on EDA than on model building. A well-understood dataset leads to better features, better models, and more trustworthy results.

Try EDA on any freely available dataset. The Government of India open data platform offers hundreds of options. Pick one, load it in Python, and start exploring today.

Technical Article

Data Collection in the Wild: Web Scraping and APIs

Data science begins with data. But where does the data come from? In textbooks, it is handed to you. In the real world, you must go and get it. Web scraping and APIs are two powerful ways to collect data at scale.

Learning these skills turns the entire internet into a potential dataset. News articles, product prices, sports statistics, and government records all become accessible with a few lines of Python code.

What Is Web Scraping?

Web scraping means automatically extracting information from websites. A Python program visits a webpage, reads its HTML structure, and pulls out the data you need. The result is a clean dataset ready for analysis.

The most popular Python library for this is BeautifulSoup. Combine it with the Requests library to fetch a webpage and parse its content. With just ten lines of code, you can extract a table from a Wikipedia article.



Fig. 4 — The web scraping pipeline: from website to stored dataset

What Is an API?

API stands for Application Programming Interface. It is a structured way for programs to exchange data with each other. Many websites offer APIs that serve data in a clean, machine-readable format called JSON.

For example, the OpenWeatherMap API gives you live weather data for any city. You send a request with a city name. The API returns temperature, humidity, and wind speed as structured data. No HTML parsing is needed.

APIs are more reliable than scraping. The data format stays consistent over time. Websites often redesign their HTML, which breaks scrapers. API formats change rarely. Always use an API when one is available.

Technical Article

Data Collection in the Wild: Web Scraping and APIs

Ethics and Legality

Web scraping raises important ethical questions. Not all websites permit scraping. Many publish rules in their robots.txt file. This file specifies which pages a bot may or may not visit. Always check it before scraping.

Sending too many requests too quickly can harm a server. It slows it down for real users. Always add a small delay between requests. The `time.sleep()` function in Python handles this easily.

Never scrape personal data without consent. Collecting names, emails, or phone numbers without permission likely violates privacy laws. Stick to publicly available, non-personal information for academic projects.

A Hands-On Ex

ercise

Try scraping a list of Indian states and capitals from a Wikipedia table. Import Requests and BeautifulSoup. Fetch the Wikipedia URL. Use BeautifulSoup to find all table rows. Loop through rows and extract cell text. Save the result to a CSV using Pandas.

This small project touches every stage of the scraping pipeline. It gives you a complete, working program you can adapt for many other tasks.

Building a Data Mindset

The best data scientists always ask: where can I find more data to test this idea? Web scraping and APIs give you the power to answer that question independently. Start small, stay ethical, and build from there.



Prof. A. K. Patil

Career Horizon

Careers in Data Science: What Roles Actually Exist?



Data science is one of the fastest-growing career fields in India. But many students are unsure what the field actually contains. They hear "data scientist" and "machine learning engineer" and wonder what the difference is. This article clears that up. The data world has several distinct roles. Each requires a different mix of skills. Understanding them helps you choose a path that fits your strengths.

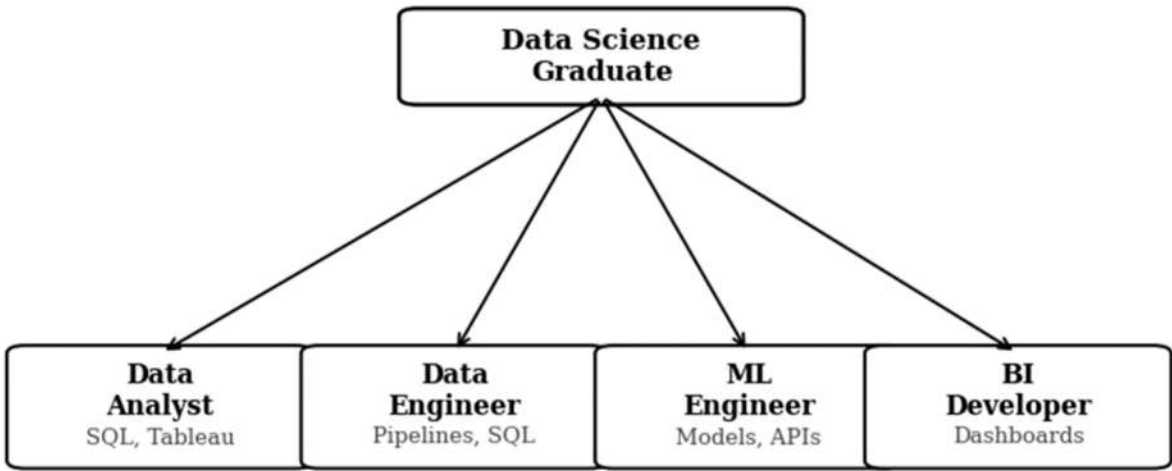


Fig. 5 — Career paths available to a Data Science graduate

Data Analyst

The data analyst is the most common entry-level role. Analysts explore data to answer specific business questions. They use SQL to query databases and Tableau to build dashboards. Strong communication skills are essential, because analysts must explain findings to non-technical teams. Typical questions a data analyst answers include: Which product category had the highest returns last quarter? Where are customers dropping off in the app?

Data Engineer

Data engineers build the pipelines that move data from one place to another. They design databases and write scripts to clean and transform raw data. Without data engineers, analysts and scientists would have no clean data to work with.

Career Horizon

Careers in Data Science: What Roles Actually Exist?

.....

This role suits students who enjoy software development and system design. Skills include SQL, Python, and tools like Apache Spark. Data engineering roles in India are growing rapidly, with strong salaries even at entry level.

Business Intelligence Developer

The BI developer creates dashboards and reports that help businesses monitor performance. They use tools like Power BI, Tableau, and Looker. Their work is highly visual. Proficiency in SQL is necessary, and familiarity with any major BI platform is a significant hiring advantage.

Machine Learning Engineer

The ML engineer takes a trained model and deploys it in a production system. They write code that serves predictions to users at low latency. Skills required include Python, cloud platforms such as AWS or Azure, and containerisation tools like Docker.

How to Get Started

The best way to enter data science is through a combination of coursework and projects. Complete a Python for data analysis course. Build three to five projects and publish them on GitHub. Participate in Kaggle competitions to develop practical skills.

Target analytics companies, e-commerce firms, and startups in fintech and healthtech for internships. These industries hire the most entry-level data professionals in India.

Data science rewards curiosity and consistency. Start building your portfolio while you are still a student. Every project you complete is evidence of real capability.



Prof. T. R. Girase