

Chapter 8: Data Science

Grade VI • Notebook Notes

Topics: Data Science • Data Science Life Cycle • Applications of Data Science

Learning Goal: Understand how data is collected, organised, analysed and used to find useful information and support decisions.

1. What is Data Science?

Data Science is a field that uses data, mathematics, statistics, computing and other methods to discover useful information, patterns and insights from data.

In simple words, **Data Science helps us turn data into useful knowledge.**

Example:

A school can study students' attendance and marks to find patterns and understand whether attendance is related to academic performance.

2. Data Science and Data

Data is the raw material used in Data Science. Data may be present as numbers, text, images, audio, video or other forms.

Data	Possible use in Data Science
Students' marks	Find average marks and identify performance patterns.
Attendance records	Identify attendance trends.
Weather readings	Study temperature or rainfall patterns.
Customer feedback	Understand opinions and common problems.

3. Why is Data Science Important?

- It helps us understand large amounts of data.
- It helps us identify patterns and trends.
- It supports better decision-making.
- It can help predict possible future outcomes.
- It helps organisations solve problems using evidence.
- It can improve services in education, healthcare and transport.

4. Data Science Life Cycle

The **Data Science Life Cycle** is a sequence of stages through which data is collected, prepared, analysed and used to obtain useful insights.

Stage	Meaning
1. Problem Definition	Clearly identify the question or problem to be solved.
2. Data Collection	Collect relevant data from suitable sources.
3. Data Preparation	Organise and clean the data by correcting errors or handling missing information.
4. Data Analysis	Study the data to find patterns, relationships and useful information.

5. Interpretation	Understand what the results mean in relation to the original problem.
6. Communication	Present findings using tables, charts, graphs or reports.
7. Action / Decision	Use the insights to make a decision or take suitable action.

5. Data Collection

Data collection means gathering the information required to answer a question or solve a problem.

- Surveys and questionnaires
- Observation
- School records
- Sensors and measuring devices
- Websites and online databases
- Applications and digital systems

Important: Data should be collected responsibly. We should respect privacy and avoid collecting personal information unnecessarily.

6. Data Preparation

Collected data may contain mistakes, duplicate entries or missing values. **Data preparation** involves organising and cleaning the data so that it can be studied properly.

- Remove or correct obvious errors.
- Organise information in a suitable format.
- Identify duplicate records.
- Handle missing information appropriately.

7. Data Analysis

Data analysis means examining data to discover useful patterns, relationships or trends.

For example, a teacher may calculate the average score of a class and compare the performance of different groups of students.

Common ways of presenting data: tables, bar graphs, line graphs, pie charts and simple summaries.

8. Applications of Data Science

Area	Example
Education	Studying learning performance, attendance and learning patterns.
Healthcare	Analysing health information to support diagnosis and treatment.
Weather	Studying weather data and making forecasts.
Transport	Analysing traffic and travel data to improve routes and services.
Sports	Studying player and team statistics.
Business	Understanding customers, sales and product demand.
Agriculture	Using weather, soil and crop data to support farming decisions.

9. Data Science in Everyday Life

- **Online shopping:** data can be used to recommend products.
- **Maps:** location and traffic data can help suggest routes.
- **Streaming services:** viewing or listening patterns can help recommend content.
- **Weather apps:** weather observations and models are used to provide forecasts.
- **Schools:** academic and attendance data can help identify learning needs.

10. Data Science and Artificial Intelligence

Data Science and Artificial Intelligence are related but are not exactly the same. Data Science focuses on extracting useful information and insights from data. AI focuses on building systems that can perform tasks associated with intelligent behaviour.

Data can be used to **train and improve many AI systems**. Therefore, good-quality data is important for many AI applications.

11. Responsible Use of Data

- Protect people's privacy.
- Use data only for appropriate purposes.
- Check data for errors.
- Avoid making unfair conclusions from incomplete data.
- Keep sensitive information secure.
- Explain important decisions when possible.

12. Quick Revision

- Data Science uses data and computational methods to discover useful insights.
- The Data Science Life Cycle begins with defining a problem and ends with communicating and using insights.
- Data must be collected carefully and prepared before analysis.
- Tables, charts and graphs can help communicate findings.
- Data Science has applications in education, healthcare, weather, transport, sports, business and agriculture.
- Responsible data use requires attention to privacy, security, accuracy and fairness.

13. Practice Questions

1. What is Data Science?
2. Why is data called the raw material of Data Science?
3. What is meant by the Data Science Life Cycle?
4. Name any four stages of the Data Science Life Cycle.
5. Why is data preparation important?
6. What is data analysis?

7. Mention any three applications of Data Science.
8. How is Data Science related to Artificial Intelligence?
9. Why should privacy be considered while collecting data?
10. Give one example of how Data Science can be useful in a school.

Notebook Tip: Draw the Data Science Life Cycle as a flow diagram: **Problem** → **Collect** → **Prepare** → **Analyse** → **Interpret** → **Communicate** → **Act**.