What is CRISP-DM?

CRISP-DM (**Cross-Industry Standard Process for Data Mining**) is a structured, industry-approved method for **data mining projects**. It provides a **step-by-step approach** to handling data to **make decisions**.

The Six Stages of CRISP-DM

CRISP-DM follows **six main steps** in a **cycle** (since data science is an iterative process).

1 Business Understanding

- The **most important stage** → Sets the **goal** of the project
- Involves talking to **stakeholders** to define **clear objectives**
- If this step is weak, the entire project could **fail**

2 Data Understanding

- Identify **data sources** and collect data
- Analyze the **quality, completeness, and relevance** of data
- Decide if more data is needed before moving forward

3 Data Preparation

- Clean and **transform raw data** into a usable format
- Handle **missing values, duplicates, and inconsistencies**
- Select relevant features for analysis

4 Modeling

- Build **machine learning or statistical models**
- Identify **patterns** in the data
- Fine-tune model parameters to improve accuracy

5 Evaluation

- Test if the model is actually **working as intended**
- Use **test data** to measure performance
- Decide if the model **needs changes** before deployment

6 Deployment

- Apply the model to **real-world data**
- Deliver insights to stakeholders for **decision-making**
- Sometimes, new data might require **model updates**

What Happens After Deployment?

- CRISP-DM is **cyclical**, meaning after deployment, you **review results** and might **redo earlier steps**.
- A **final business meeting** with stakeholders ensures that the model actually helps in decision-making.

Key Takeaways:

✅ **CRISP-DM is an iterative process** → You **revisit steps** when necessary.
✅ **The project must align with business goals** → Otherwise, it's a waste of time.
✅ **Evaluation and deployment aren't the end** → You might need to **adjust and improve**.

CRISP-DM is a **specific** framework for **data mining**, while **Data Science Methodology** is a **broader** concept that includes **different approaches** (one of which is CRISP-DM).

Key Differences:

Feature	CRISP-DM	Data Science Methodology
Focus	Data **mining** projects	**General** data science process
Structure	**6-step** process (Business Understanding → Deployment)	More **flexible** with different frameworks (e.g., CRISP-DM, KDD, SEMMA)
Iteration	Explicitly **cyclical** (stages repeat if needed)	Also **iterative**, but depends on the chosen framework
Use Case	Best for structured **business problems**	Covers **all** aspects of data science (data mining, AI, ML, etc.)

Are They Related?

Yes! **CRISP-DM** is one way to follow a **data science methodology**, but it's not the **only** one. If you're doing **machine learning, AI, or advanced analytics**, other frameworks might fit better.

So, **CRISP-DM is a part of Data Science Methodology**, but **not the same as** the entire field. 🚀