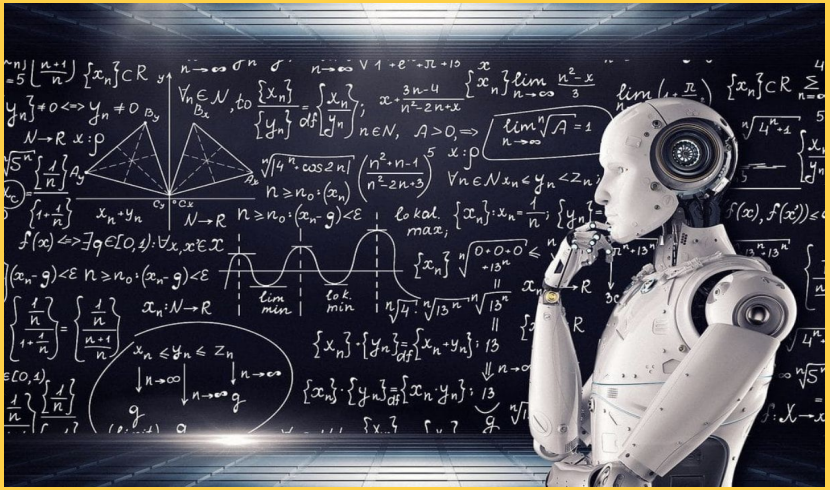
Introduction To Data Science

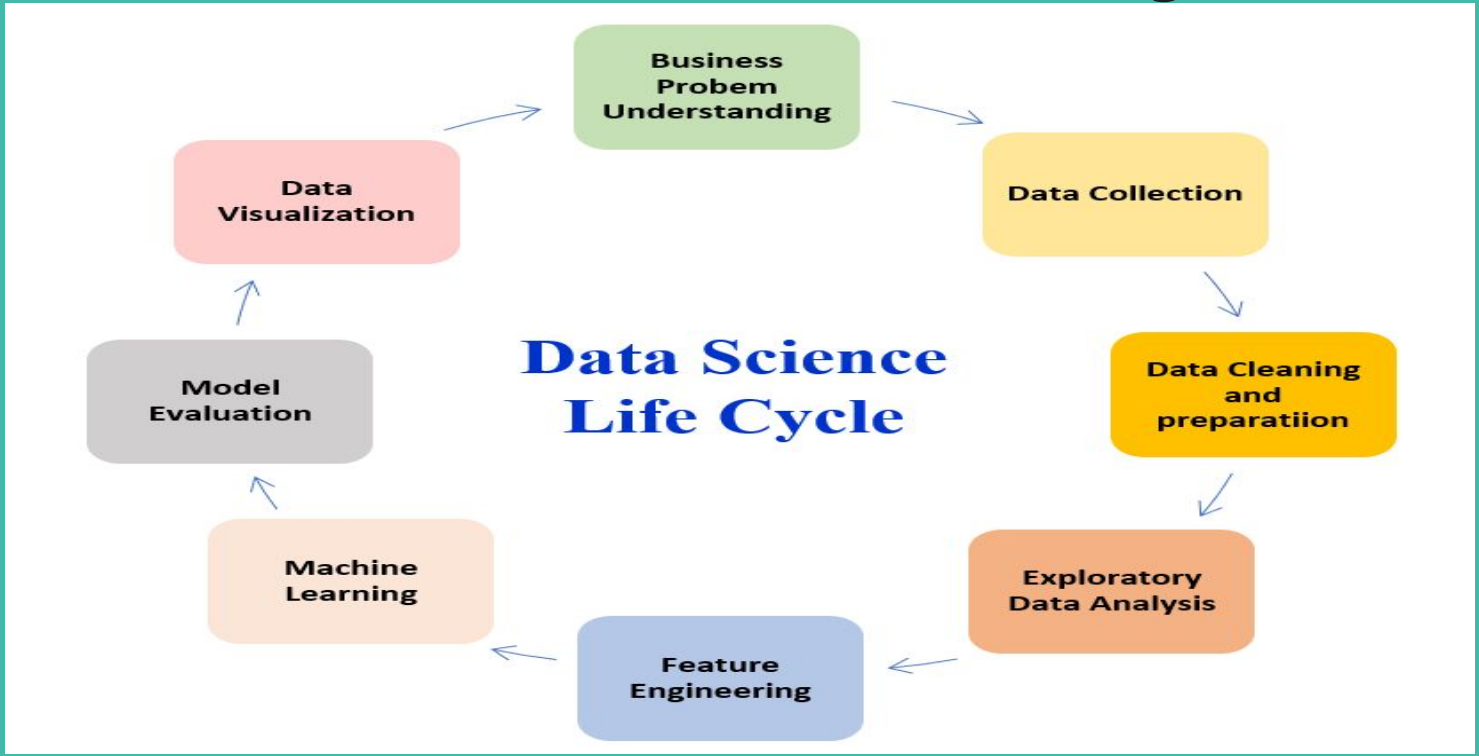
Machine Learning and Data Science Club x ColorStack

What is Data Science?

- Data science is the field focused on transforming raw data into meaningful insights by designing processes, systems, and tools that support data-driven decision making across many industries.
- It combines multiple skill sets: including statistics, data visualization, programming, data wrangling, databases (SQL/NoSQL), and machine learning to understand patterns and make predictions.
- Its goal is to turn data into action, helping organizations explain what happened, understand why it happened, and inform what to do next.



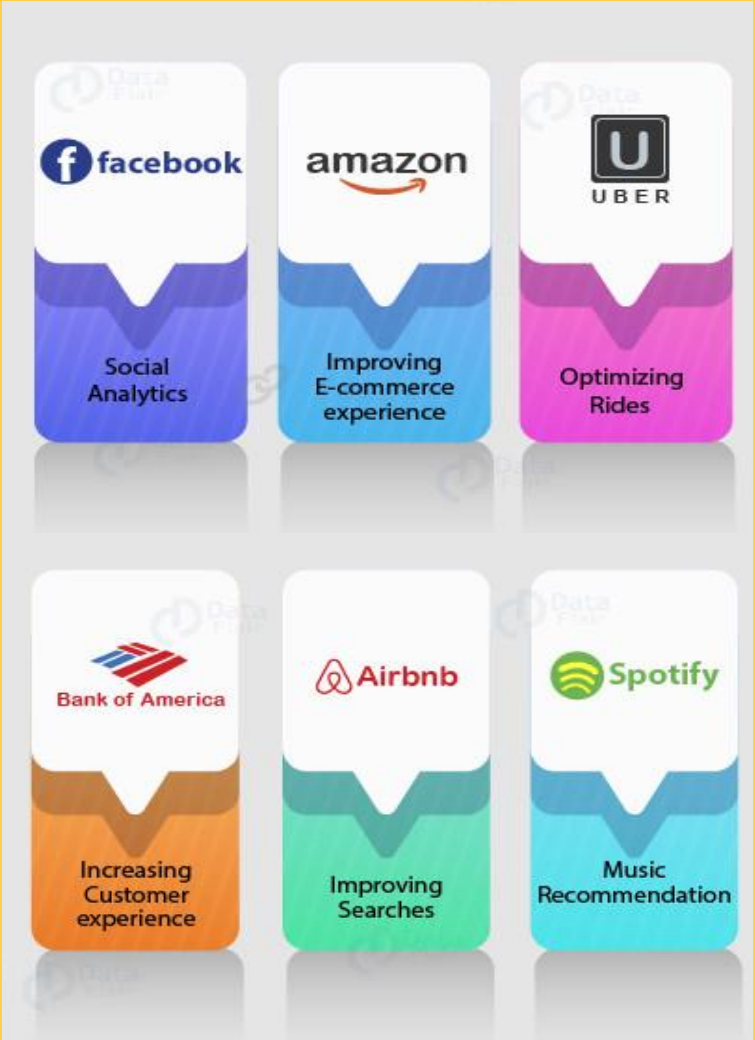
The Data Science Life Cycle



Where Data Science Is Used

Data science is used across many industries to analyze data, uncover patterns, and support better decision-making.

- **Social media**: content recommendations, trend detection
- **Finance**: risk analysis, fraud detection, forecasting
- **Healthcare**: patient monitoring, diagnosis support
- **Marketing**: customer segmentation, targeted advertising
- **Sports** & **education**: performance analysis, outcome prediction



Types of Data & Tools

Data comes in different forms, and understanding these differences helps determine how it can be analyzed.

- Structured data: organized in rows and columns (tables, spreadsheets)
- Unstructured data: text, images, audio, video
- Numerical data: values that can be measured or counted
- Categorical data: labels or groups (e.g., categories, types)

Common tools used in data science include:

- Programming languages: Python, R
- Libraries: pandas (data handling), matplotlib (visualization), scikit-learn (machine learning)
- Databases: SQL and NoSQL for storing and querying data

From Data to Decisions

The goal of data science is not just to build models, but to turn data into insights that inform real-world decisions.

- **Insights should be clear, interpretable, and actionable**
- **Good decisions depend on data quality, not just model complexity**
- **Models have assumptions and limitations that must be understood**
- **Results must be communicated effectively using visuals and storytelling**

Next, we'll apply this process hands-on using a real dataset.