

Data Science

The field

Definition:

Data Science is the field of using data to find patterns, knowledge, and solutions.

Example:

Netflix uses data science to recommend movies and shows.

It is the field where Netflix collects users' data (which movies they watch, how long they watch, which actors they like) and finds patterns to predict what new movies a user may like.

Data Scientist

The expert(person)

Definition:

A Data Scientist is a person who collects, cleans, and analyzes data to make predictions and better decisions.

Example:

He/She is the expert at Netflix who collects this data, cleans it, builds algorithms, and finally creates the recommendation system that suggests movies to you.



What a Data Scientist Can Do

A Data Scientist is like a **problem-solver** who uses data to **understand the past, explain the present, and predict the future**. A Data Scientist works with data to help companies make better decisions. Here are the main things they can do:

1. **Collect Data** → Take data from websites, apps, sensors, databases.
 - Example: Netflix collects data about what users watch.
2. **Clean Data** → Fix errors, remove duplicates, handle missing values.
 - Example: Removing wrong or incomplete customer records.
3. **Analyze Data** → Use math and statistics to find patterns.
 - Example: Finding which product sells more in summer vs winter.
4. **Build Models** → Create machine learning models for predictions.
 - Example: Predicting which customers might stop using a service.
5. **Visualize Data** → Make charts, graphs, dashboards to explain results.
 - Example: A graph showing sales growth of a company.
6. **Make Predictions** → Forecast future trends using past data.
 - Example: Weather apps predicting tomorrow's rain.

7. **Help in Decision-Making** → Give insights to businesses.
 - Example: A bank deciding loan approval based on data analysis.

Main and Important Jobs of Data Scientist:

1. **Data Scientist** → Analyze big data & find insights
2. **Data Analyst** → Make reports & visualize data
3. **Machine Learning Engineer** → Build AI models
4. **Data Engineer** → Manage & clean data pipelines
5. **Business Intelligence Analyst** → Help businesses make data-based decisions

Main Work: Collect data → Clean it → Analyze → Build models → Give insights for decision making.

Top 5 Countries for Data Scientist Jobs

1. **USA US** → Most job opportunities
 2. **UK GB** → Good salary and opportunities
 3. **Switzerland CH** → Highest salaries
 4. **India IN** → Fast-growing job market
 5. **Singapore SG** → Best for AI and Data Science
-

Main Companies Hiring Data Scientists

- **Tech:** Google, Amazon, Microsoft
 - **Finance:** JPMorgan, Goldman Sachs
 - **Healthcare:** Philips, Medtronic
 - **Research/Education:** MIT, Stanford
 - **Government:** NASA, ISRO
-

IDE?

 **IDE (Integrated Development Environment)** is a software where you can **write, run, and test Python code easily.**

Top 5 IDEs for Python

1. **PyCharm** → Best for professional projects.
2. **VS Code** → Lightweight, popular, and supports many languages.
3. **Jupyter Notebook** → Great for data science and machine learning.
4. **Spyder** → Good for scientific programming.

5. **Thonny** → Simple and easy for beginners.

Top Python VS Code Extensions

1. Python
2. python test explorer
3. Python Indent
4. Python extension pack
5. AREPL for Python
6. Black Formatter
7. better Comments
8. Intellicode
9. Jupyter /jupyter keymap
10. Codesnap
11. Markdown all in one
12. Prettier
13. Blackbox
14. Pylance
15. autoDocstring

keywords in python:

Keywords are special **reserved words** in Python

- They have **fixed meaning**.
- You **cannot use them** as variable names or identifiers.
- Python keywords are **case-sensitive** → True , true
- You **cannot use them** as variable names.
- They are used to **control flow, logic, and structure** of the program.
- Python has **reserved words** (cannot be changed).
- They are all **lowercase** (mostly).
- They control the **structure** of a program.
- Number of keywords may change with Python version.
- If you try to use them as variable names → Error.

```
False    await    else    import    pass  
None    break    except    in     raise  
True    class    finally    is      return  
and     continue  for     lambda   try
```

```
as      def      from      nonlocal  while
assert  del      global    not      with
async   elif     if       or       yield
```