🧩 Data Exploration: Solving the Data Puzzle

🎯 What is Data Exploration?

- The process of **understanding a dataset** by:
  - Inspecting its characteristics
  - Identifying patterns
  - Asking questions

🧠 Analogy

- Like solving a **jigsaw puzzle**: you examine each piece to see how it fits into the bigger picture.

🔍 Techniques Used in Data Exploration

1. Sampling

- Selecting a **representative segment** of the dataset.
- Helps analysts work with a **manageable subset** to understand the data's structure and ask relevant questions.

2. Visual Exploration

- Using **histograms and scatter plots** to:
  - Spot patterns
  - Identify **outliers and anomalies**
  - Improve **data quality** before analysis

3. Data Drilling

- Exploring data at a **granular level**.
- Helps refine questions and build **specific queries** for deeper insights.

👤 Example: Arjun the Analyst

- Works at a large clothing company.
- Uses **sampling**, **visualization**, and **data drilling** to understand a massive dataset.
- These techniques help him **see the big picture** and prepare for meaningful analysis.

🔁 Important Takeaway

- **Data exploration is iterative**, not linear.
  - You may need to **go back and forth** between steps.
  - Your approach will vary depending on the **data and goals**.

✅ Why It Matters

- Helps you:
  - Understand data better
  - Save time and resources
  - Ensure your analysis is **accurate, insightful, and aligned** with your objectives