Data Science Fundamentals: Essential Notes

1. Introduction to Data Science

- We live in a world surrounded by data, generated constantly through digital interactions (e.g., social media, e-commerce, banking).

- **Comment:** Understanding this context shows how central data is to modern life and how it drives change in organizations.

---

2. What is Data?

- **Data** is simply a collection of facts, which could be:

    - Numbers

    - Alphabets

    - Alphanumeric strings

    - Symbols

    - Images, Audio, Video, etc.

- **Comment:** Think of data as the raw material for all analysis—it's unprocessed and lacks context.1.-DATA-SCIENCE-FUNDAMENTALS.PDF

---

3. What is Information?

- **Information** is processed data—the result of organizing and interpreting data so that it becomes meaningful and useful.

- **Comment:** A crucial skill for any data scientist is the ability to transform data into valuable information through contextualization.1.-DATA-SCIENCE-FUNDAMENTALS.PDF

---

4. Data Processing

- **Data Processing** is the act of converting raw data into meaningful information.

- Example: Names, professions, and locations as raw data can be organized into structured records (see example tables in source).1.-DATA-SCIENCE-FUNDAMENTALS.PDF

---

5. What is Data Science?

- **Data Science** uses scientific methods and algorithms to extract knowledge and insights from data.

- Includes steps like:

  - Processing data

  - Understanding historical and current trends

  - Supporting decision-making

  - Predicting future outcomes

  - Creating new solutions and products

- **Comment:** Data science overlaps with fields like statistics, computer science, business, and domain expertise—communication is key for impact.1.-DATA-SCIENCE-FUNDAMENTALS.PDF

---

6. Who is a Scientist? (in Data Science)

- A scientist systematically gathers information, performs research, tests evidence, and shares knowledge.

- **Comment:** Cultivating curiosity and critical thinking is vital for success in data science.

---

7. Big Data

- **Big Data** refers to data sets so large and complex that traditional processing tools cannot handle them.

- Examples: Social media streams, sensor data from IoT, transaction logs.1.-DATA-SCIENCE-FUNDAMENTALS.PDF

---

8. Types of Big Data

Type	Description	Examples
Structured	Table formats (rows/columns). Easy to analyze.	Database tables
Semi-structured	Partial structure (tags/attributes) but not fully tabular	XML, JSON files
Unstructured	No predefined structure	Emails, videos

-
  Comment: Most real-world data is unstructured, posing unique challenges for storage and analysis.1.-DATA-SCIENCE-FUNDAMENTALS.PDF

---

9. 3 Vs of Big Data

- **Volume:** Sheer size of data generated and stored.

- **Variety:** Different formats and types (structured, semi-structured, unstructured).
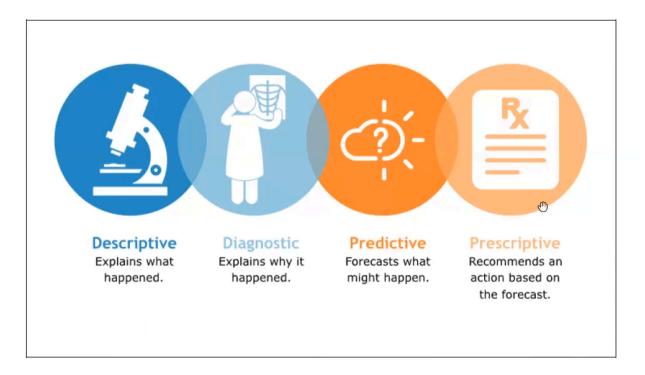
- **Velocity:** Speed at which new data is generated and processed.

- **Comment:** Sometimes, more Vs are mentioned (Veracity, Value)—being aware of data quality and usefulness is important.

---

10. Data Analytics

- Analysis of data (especially big data) to improve decision making, customer understanding, and business performance.

- **Comment:** Every productive company leverages analytics to stay competitive and innovative.1.-DATA-SCIENCE-FUNDAMENTALS.PDF

---

11. Types of Analytics

Type	Purpose	Example
Descriptive	What happened?	Sales trends last quarter
Diagnostic	Why did it happen?	Root cause analysis of a sales drop
Predictive	What will happen?	Next month's sales forecast
Prescriptive	What should we do next?	Recommended marketing strategies

-

Additional Comments & Tips

- **Learn essential tools:** Python/R, SQL, spreadsheets, data visualization tools (e.g., Tableau, Power BI).

- **Sharpen critical thinking:** Always ask what business or research question you are answering.

- **Stay ethical:** Handle data privacy and bias responsibly.

- **Practice communication:** Explaining findings to both technical and non-technical audiences is indispensable.

- **Build a portfolio:** Solve real problems and document your work on GitHub or personal blogs.