**Theoretical Assignments:**

1. **Write an essay: Importance of Data Cleaning in Data Science.**

* **Importance of Data Cleaning in Data Science**

the ability to derive meaningful insights from massive datasets is central to data science.

However, raw data is often messy, incomplete, inconsistent, and riddled with errors.

This is where **data cleaning** becomes essential.

1. Accuracy and Quality

The primary purpose of data cleaning is to improve the accuracy and quality of the data.

2.Improves Model Performance

the quality of the input data directly affects the performance of the model.

3.Reduces Processing Time

allowing for more efficient data processing and faster execution of analytical workflows.

4. Supports Compliance and Data Governance

Many industries, such as finance, healthcare, and e-commerce, are subject to strict data governance and regulatory requirements.

5. Enhances Data Visualization

**2.Presentation: Data Visualization Techniques & Best Practices**

**What is Data Visualization?**

* Definition: The graphical representation of information and data.
* Purpose: To help users see and understand patterns, trends, and outliers in data.

**Why is Data Visualization Important?**

* Enhanced Understanding: Makes complex data digestible and comprehensible.
* Faster Insights: Quickly identify trends, correlations, and anomalies.
* Improved Decision-Making: Supports informed and data-driven choices.
* Effective Communication: Conveys information clearly and persuasively to diverse audiences.

**Common Data Visualization Techniques:-**

**Bar Charts:**

* Purpose: Comparing discrete categories.
* Example: Vertical vs. Horizontal bars.

**Line Charts:**

* Purpose: Showing trends over time.
* Example: Single or multiple lines.

**Pie Charts:**

* Purpose: Showing parts of a whole (proportions).
* When to Use: Market share, budget allocation.

**Scatter Plots:**

* Purpose: Showing the relationship between two numerical variables.
* Example: Identifying clusters or outliers.

**Histograms:**

* Purpose: Displaying the distribution of a single numerical variable.
* Example: Understanding data shape .