Phase 2

Theoretical Tasks

1. End-to-End Data Analysis Lifecycle (Business Example)

The data analysis lifecycle starts with understanding the business problem. For example, a retail company wants to increase its sales revenue. First, the analyst collects data such as sales, customer details, product categories, and regions. Next, the data is cleaned by handling missing values, removing duplicates, and correcting errors.

After cleaning, the analyst explores the data to find trends, patterns, and relationships. Visualization and statistical techniques are used to understand performance. Then, models or summaries are built to answer business questions like which products sell the most or which region generates more profit. Finally, insights are communicated using dashboards and reports so business users can take decisions such as increasing marketing in low-performing regions.

---

2. Descriptive vs Diagnostic Analytics

Descriptive analytics explains what happened in the past. It summarizes data using averages, totals, charts, and KPIs. Example: total sales in 2024.

Diagnostic analytics explains why something happened. It goes deeper using comparisons, drill-downs, and correlations. Example: sales dropped because customer returns increased in one region.

---

3. Correlation vs Causation

Correlation means two variables move together, but one may not cause the other. For example, ice cream sales and temperature increase together.

Causation means one variable directly causes another. For example, heavy discounts cause an increase in sales.

Correlation does not always imply causation.

---

4. Short Notes

a. Data Bias

Data bias happens when data does not represent reality correctly. It can come from missing groups, wrong sampling, or human errors. Biased data leads to wrong business decisions.

b. Missing Data Strategies

Missing data can be handled by deleting rows, filling with mean/median, using forward fill, or predicting values using models depending on the business case.

c. KPIs vs Metrics

Metrics measure activities like number of orders. KPIs measure performance linked to goals like revenue growth or customer retention.

---

5. Case:

Why Do Dashboards Fail Even With Correct Data?

Dashboards fail when they are too complex, poorly designed, or not aligned with business needs. Sometimes they show too many charts without explaining insights. Another reason is lack of interaction like filters. Also, if stakeholders don’t understand the visuals, the dashboard becomes useless even if data is correct.