

Criteria	Marks	Description
1. Understanding of the Dataset	15	<ul style="list-style-type: none"> 0–5: Little to no understanding of the dataset's purpose, variables, or context. 6–10: Reasonable understanding, some mention of variable types or purpose. 11–15: Clear, thorough explanation of the dataset, variables, and potential issues.
2. Data Cleaning	50	
Handling Outliers and Inconsistencies	15	<ul style="list-style-type: none"> 0–5: Not identified or incorrectly treated. 6–10: Some effort to manage outliers/inconsistencies. 11–15: Proper detection and treatment with clear justification.
handling Missing Values	20	<ul style="list-style-type: none"> 0–5: Missing values ignored or poorly handled. 6–10: Basic handling (e.g., deletion or simple imputation). 11–15: Thoughtful handling based on variable type or context. 16–20: Creative or well-justified approaches (e.g., separate category, domain-specific logic).
Documentation of Cleaning Steps	15	<ul style="list-style-type: none"> 0–5: Little to no explanation. 6–10: Basic comments or notes. 11–15: Clear, step-by-step documentation of all cleaning decisions.
3. Exploratory Data Analysis	20	
Descriptive Statistics	10	<ul style="list-style-type: none"> 0–4: Minimal or irrelevant statistics. 5–7: Useful summaries for most variables (mean, median, mode, etc.). 8–10: Strong statistical insight with thoughtful interpretations.
Insights Derived	10	<ul style="list-style-type: none"> 0–4: Limited or obvious insights. 5–7: Some meaningful patterns or anomalies discussed

		<ul style="list-style-type: none"> • 8–10: Deep understanding shown through key observations and data trends
4. Code Quality and Reproducibility	10	<ul style="list-style-type: none"> • 0-3: Unclear or unstructured code. • 4–6: Readable and mostly organized code. • 7-10: Clean, well-documented, and reproducible script or notebook.