**Assignment 1 – Enrin Debbarma**

1. Dataset given: Hw\_1\_data1.csv
2. 1. Data Cleaning

2. Data Integration

3. Data Reduction

4. Data Transformation

ii. 1. Data Cleaning: Check each feature for missing values, outliers, or impossible values and then correct or remove them as needed.

2. Data Integration: Ensure that the IDs match across tables and that measurement units are consistent.

3. Data Reduction: If certain features are redundant or highly correlated, I would either drop them or perform dimensionality reduction.

4. Data Transformation: Encoding - Since Feature D and Feature E are categorical, we could label‐encode or one‐hot‐encode them to use in many machine‐learning algorithms. Normalization or standardization: We could normalize or standardize columns like “Feature A,” “Feature B,” “Feature C,” and any other numeric columns if you are using algorithms sensitive to feature scale.