

5.1 Explain the Missing Values Solutions Used in Data Cleaning Process:

1. Filling in Missing Values
2. Dropping Missing Values

5.2 What Meant by Ready-to-Analyze Format?

You need to convert data into other formats by transforming data points or by subset data using some arbitrary or pre-defined conditions.

5.3 Missing Values Consider a Problem When Cleaning Genomics Data, Explain:

Most algorithms cannot deal with the missing values data because this may cause incorrect results or biases in results, low efficiency.

5.4 What Are Reasons of Imperfect Data?

Because of combining data from multiple sources, incorrect measurements, noisy and missing values.

5.5 Define the Following Terms:

1. **Genome:**
The full DNA sequence of organism that contains the heredity information.
2. **Gene:**
It's the functional unit of DNA, segments that're correspond to protein. Catalytic or structural RNA.
3. **Data Analysis:**
The process of organizing, storing, and displaying data to answer question of interest "research".
4. **Data Variable Re-coding:**
It means using different values for one variable.
5. **Data Transformation:**
It means converting data from one format into another.
6. **Data Structural Errors:**
When you measure or transfer data and notice strange naming conventions, typos, or incorrect capitalization.
7. **Data Visualization:**
It means using figures, tables, Numbers to represent the outcome of your research.

5.6 What Are Data Quality Characteristics?

Accuracy, availability, validity, accuracy, uniformity and reliability.