



PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE19CS203 – STATISTICS FOR DATA SCIENCE

Unit - 1 - Introduction to Data Science

SOLVED QUESTION BANK

Data Cleaning

1. Define Data Cleaning.

Data cleansing or **data cleaning** is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database. It identifies incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.

2. What is called as Imputation? Mention its types.

Imputation is the process of replacing missing data with substituted values.

Types

1) Single imputation.

- Hot Deck
- Cold Deck
- Mean Substitution
- Regression

2) Multiple imputation.