EXPERIMENT NO: 4

Mapped Course Outcomes- CO2

Aim:

Handling Missing Values using IBM SPSS Modeler

Objective:

We need to handle missing values in the dataset

1) Deleting the missing values using Select Node of IBM SPSS Modeler
2) Imputing the missing values using Missing Value Imputation Super node of IBM SPSS Modeler

Related Theory:

1. What are missing values
2. How to handle missing values: Numeric, Categorical etc.
3. What is Select Node
4. What is SuperNode

Procedure:

Deleting the missing values using Select Node of IBM SPSS Modeler.

1) Import Dataset
2) Connect with type node
3) With Data audit node check for Null values present
4) Use select node to filter the data with Null Values
5) Use Data Audit node to verify whether the null values has been filtered or not

Imputing the missing values using Missing Value Imputation Super node of IBM SPSS Modeler

1) Import Dataset
2) Connect with type node
3) With Data audit node check for Null values present
4) In Quality Tab of Data Audit Node, select Impute Missing dropdown as "Specify" and specify the conditions of imputing the missing values.
5) Generate Missing Value SuperNode
6) Use Data Audit node to verify whether the null values has been imputed or not.
7) Connect Missing Value SuperNode with Table node to verify the values.

Output:

Missing values are imputed in the dataset

Viva Questions

1. Which node of IBM SPSS Modeler can be used to impute (estimate) missing values?
2. Which node correct imbalances using specified condition in SPSS Modeler
3. ___ palette tab contains nodes that you can use to perform operations on the data records, such as selecting, merging, and appending.'

4. _____ perform operations such as filtering, deriving new fields, and determining the
measurement level for given fields.

5. _____ produce various output that can be viewed in external applications, such as IBM® SPSS Data Collection or Excel.