

Here's a step-by-step summary of how I completed the assignment on data pre-processing using the "BIG MART SALES DATASET" from Kaggle:

### 1. Data Loading:

- You started by importing the pandas library as ``pd``.
- Loaded the dataset using ``pd.read_csv("Test.csv")`` and stored it in the variable ``dataset``.
- Displayed the dataset using ``dataset``.

### 2. Handling Missing Values:

- Checked for missing values using both ``dataset.isnull().sum()`` and ``dataset.isna().sum()``.
- Identified two columns with missing values: "Item\_Weight" and "Outlet\_Size".
- Filled missing values in the "Item\_Weight" and "Outlet\_Size" columns with zeros using ``fillna()`` method.

### 3. Distinguishing between Quantitative and Qualitative Columns:

- Defined a function ``quanQual(dataset)`` to categorize columns into quantitative and qualitative types.
- In the function, iterated through columns and classified them based on their data types.
- Returned two lists: ``quan`` containing quantitative column names and ``qual`` containing qualitative column names.

### 4. Handling Quantitative Columns with Missing Values:

- Retrieved quantitative and qualitative column names using ``quanQual(dataset)``.
- Used the ``SimpleImputer`` class from ``sklearn.impute`` to handle missing values in quantitative columns.
- Created an instance of ``SimpleImputer`` with ``strategy='mean'`` to replace missing values with the mean of each column.
- Fitted the imputer on quantitative columns using ``imp.fit(dataset[quan])``.
- Transformed the dataset using ``imp.transform(dataset[quan])`` to replace missing values with the calculated means.
- Converted the transformed array back to a DataFrame with appropriate column names using ``pd.DataFrame(df, columns=quan)``.

#### 5. Finalizing Pre-processed Dataset:

- The pre-processed dataset with missing values handled and imputed with mean values for quantitative columns was stored in the variable `df`.

By following these steps, I successfully completed the assignment on data pre-processing without missing values in the "BIG MART SALES DATASET" from Kaggle. This process ensured that the dataset was cleaned and ready for further analysis or modelling.