Day – 4 Report

Task – The R Language Challenge

Day – Monday

Date – 16th Feb 2026

Time - 7:00 AM – 10:00 AM

Dataset - iris_flower.

Source – The dataset was downloaded from the Kaggle.

- This dataset is about the **iris flower**. It is a biological dataset which contains observations of iris flower.
- Firstly, analyzed the dataset using **MS-Excel.**
- It has **150 Rows** and **5 Columns.**
- Where the columns represent the **physical measurements** of iris flower.
- Column are sepal_length_cm, sepal_width_cm, petal_length_cm, petal_width_cm, class.
- The dataset was imported into the **RStudio** for further analysis and cleaning purpose.
- Observed the dataset structure.
- Checked for the **missing** and **duplicated** values.
- As result there are **5 missing values** in dataset in column **petal_width_cm,** and **no duplicated** values were found.
- Handled missing values by using **median operation** on the column because the data was small and median will be more effective to maintain the data.
- After performing all operations and cleaning the dataset saved the cleaned dataset.