

Iris Dataset Cleaning Report.

For today's task, the Iris dataset was used.

The Iris dataset contains **150 rows** and **5 columns**. The column names are **sepal length**, **sepal width**, **petal length**, **petal width**, and **species**.

The dataset was cleaned using the R programming language in RStudio. First, the required libraries were loaded to help in reading and cleaning the data.

The dataset was imported into R using the `read_csv()` function. After loading the data, the size of the dataset was checked using the `dim()` function. The column names and data types were.

Next, the dataset was checked for missing values using the `is.na()` function. The results showed that there were no missing values in the Iris dataset. The dataset was also checked for duplicate rows, there were duplicate records. So using `distinct` function duplicate rows were deleted.

For **sepal length**, **sepal width**, **petal length** and **petal width**, **median** values used to fill the missing values.

Then for species, Condition was added to classify species using petal length. Flowers with short petal length were classified as **setosa**, flowers with medium petal length were classified as **versicolor**, and flowers with long petal length were classified as **virginica**.

Finally, the dataset was checked again to make sure everything was correct. After confirming the data was clean, it was saved as a new CSV file using the `write_csv()` function.