**Assignment No:-1**

**Problem Statement:-**

Perform the following operations using R/Python on suitable data sets:

a) read data from different formats (like csv, xls)

b) indexing and selecting data, sort data,

c) describe attributes of data, checking data types of each column,

d) counting unique values of data, format of each column, converting variable data type

(e.g. from long to short, vice versa),

e) identifying missing values and fill in the missing values

**Theory**:-

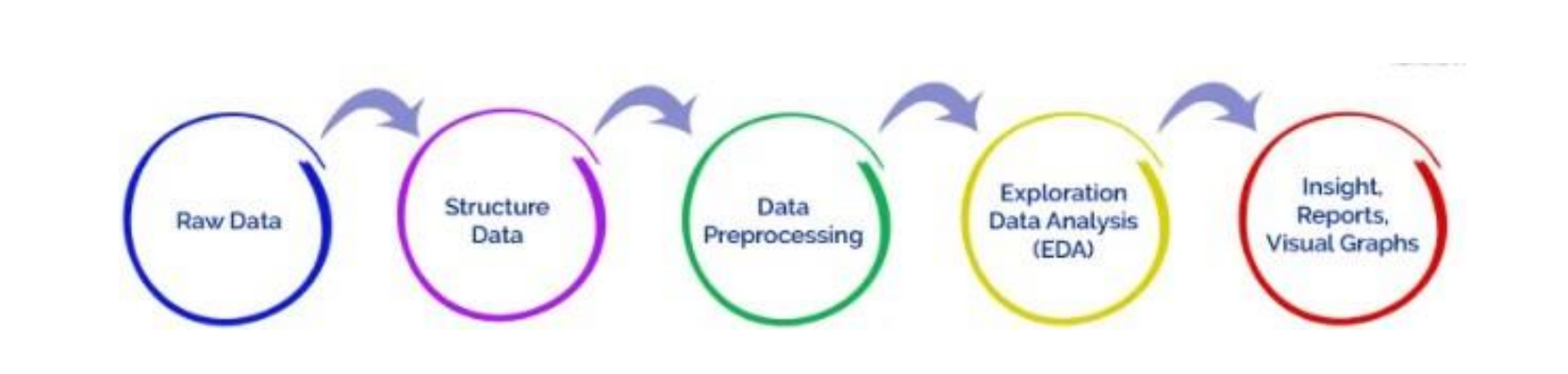
**Methodology:**

1. Reading Data from Different Formats:
   * For reading data from CSV files, you can use read\_csv() function in Python's pandas library or read.csv() function in R.
   * For reading data from Excel files, you can use read\_excel() function in pandas or read\_excel() function in R.
2. Indexing and Selecting Data, Sorting Data:
   * Indexing and selecting data can be done using square brackets [] in both Python (pandas) and R.
   * For sorting data, you can use sort\_values() function in pandas and arrange() function in dplyr package in R.
3. Describing Attributes of Data, Checking Data Types:
   * You can use info() function in pandas and str() function in R to describe attributes of data and check data types of each column.
4. Counting Unique Values, Converting Variable Data Type:
   * To count unique values of data, you can use value\_counts() function in pandas and table() function in R.
   * For converting variable data type, you can use astype() function in pandas and as.type() function in R.
5. Identifying Missing Values and Filling in Missing Values:
   * To identify missing values, you can use isnull() function in pandas and is.na() function in R.
   * To fill in missing values, you can use fillna() function in pandas and na.omit() function in R.

Advantages and Disadvantages & Limitation/Example:

1. Advantages:
   * Both R and Python offer rich libraries (pandas in Python, tidyverse in R) for data manipulation tasks.
   * Flexibility in terms of handling various data formats.
   * High performance and efficiency for handling large datasets.
   * Extensive documentation and community support for both languages.
2. Disadvantages & Limitations/Example:
   * *Learning Curve*: Both R and Python have a learning curve, especially for beginners with no prior programming experience.
   * *Package Compatibility*: Occasionally, certain packages might not be available for the latest version of R/Python or might not be compatible with each other.
   * *Syntax Differences*: Syntax differences between R and Python might lead to confusion for users switching between the two.
   * *Performance*: Depending on the task and implementation, one language might be more efficient than the other. For instance, R might be slower than Python for certain types of tasks, especially those involving loops.

**Diagram**



**Conclusion**

We were able to perform reading,indexing,counting and many more such operatons on datasets in python language.