**Problem Statement**

The task involves performing various data manipulation operations using R/Python on suitable datasets. These operations include reading data from different formats, indexing and selecting data, sorting data, describing attributes of data, checking data types of each column, counting unique values of data, formatting each column, converting variable data type, identifying missing values, and filling in the missing values.

**Libraries Used**

- pandas

- numpy

- matplotlib

- seaborn

- plotly

- scikit-learn

**Theory**

\*\*Methodology:\*\* The methodology involves using various functions and methods provided by the libraries to perform the specified data manipulation tasks. These tasks are essential for data preprocessing and exploratory data analysis, which are crucial steps in the machine learning pipeline.

\*\*Advantages:\*\*

- Enables efficient data handling and manipulation.

- Facilitates data exploration and understanding.

- Helps in preparing data for machine learning algorithms.

\*\*Applications:\*\*

- Data preprocessing for machine learning tasks.

- Exploratory data analysis to gain insights into the dataset.

- Data cleaning and transformation for better model performance.

\*\*Limitations:\*\*

- May require domain knowledge to interpret results accurately.

- Data manipulation tasks can be time-consuming for large datasets.

- Errors in data handling can lead to incorrect analysis and modeling results.

**Working Algorithm**

1. \*\*Read Data:\*\* Use appropriate functions to read data from different formats such as CSV and XLS.

2. \*\*Indexing and Selection:\*\* Use indexing and selection techniques to extract specific rows and columns from the dataset.

3. \*\*Sorting Data:\*\* Sort the data based on one or more columns to arrange it in a particular order.

4. \*\*Describing Attributes:\*\* Use descriptive statistics functions to describe attributes of data, including mean, median, standard deviation, etc.

5. \*\*Checking Data Types:\*\* Check the data types of each column in the dataset to ensure consistency and compatibility with analysis methods.

6. \*\*Counting Unique Values:\*\* Count the number of unique values in each column to understand the data distribution.

7. \*\*Converting Data Types:\*\* Convert variable data types as needed, such as converting from long to short format or vice versa.

8. \*\*Identifying Missing Values:\*\* Identify missing values in the dataset using functions like isnull() or isna().

9. \*\*Filling Missing Values:\*\* Fill in the missing values using appropriate techniques such as mean imputation, mode imputation, or using machine learning models.

**Conclusion**

In this project, we performed various data manipulation operations using R/Python on suitable datasets. These operations are fundamental for data preprocessing and exploratory data analysis, which are essential steps in the machine learning pipeline. By understanding and applying these operations effectively, we can prepare the data for modeling and gain valuable insights for further analysis.