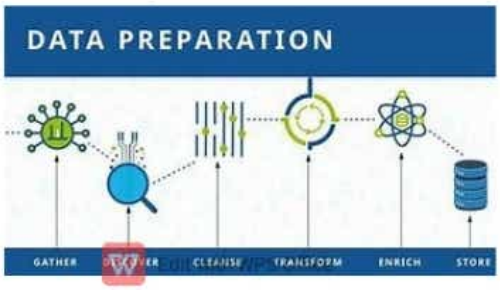
# MEASURE ENERGY CONSUMPTION

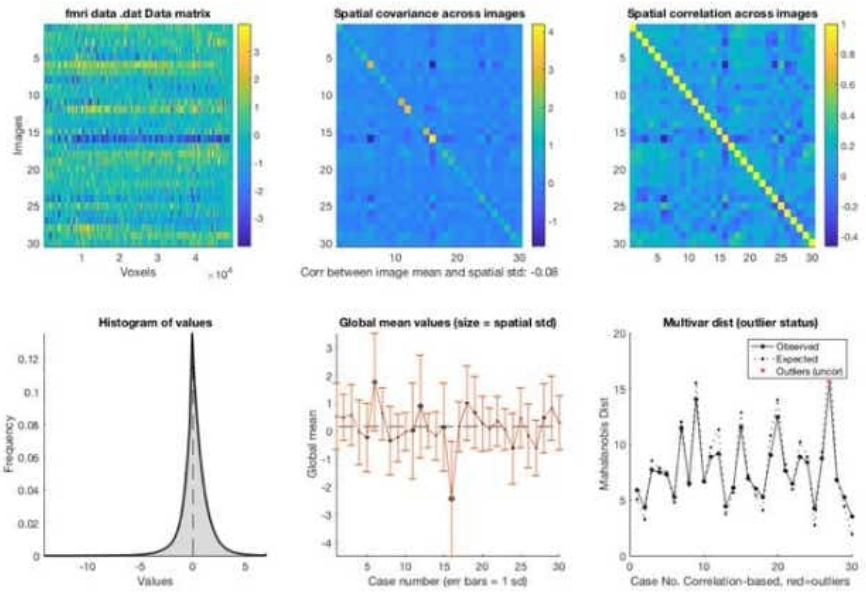
## TEAM MEMBER

510521104701 SANTHOSH V

Phase 3: DOCUMENT SUBMISSION

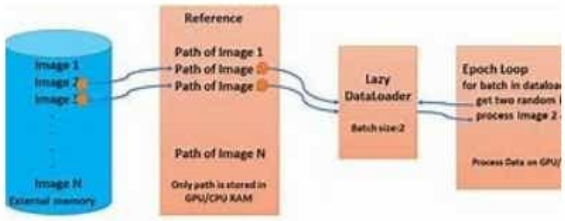
Project Title : LOADING AND PREPROCESSING THE DATASET





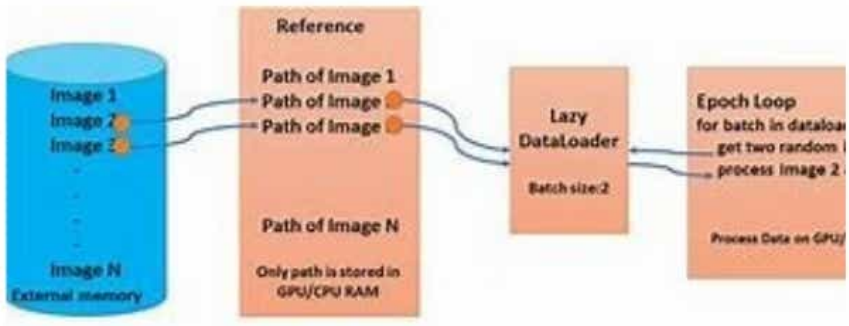
Introduction :

"Loading and preprocessing the dataset is a fundamental step in data analysis and machine learning. This crucial process involves acquiring and structuring data, handling missing values, cleaning, and feature engineering, ensuring the dataset is ready for analysis or model training. A well-executed load and preprocessing phase lays the foundation for robust insights and accurate predictions."

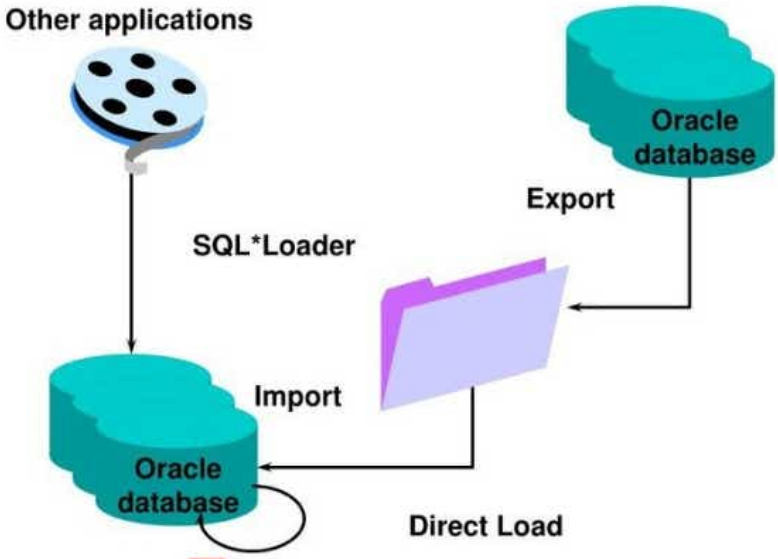


Data Loading :

Explain the process of loading a dataset into your chosen tool or programming language (e.g., Python, R).
You can include a code snippet or screenshots of loading data.



# Data Loading Methods

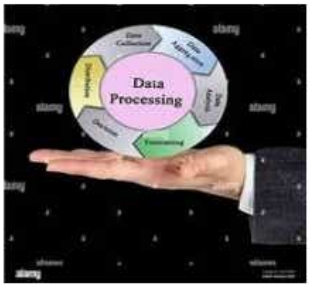


Putting all of this together, we can now load the data and summarize the loaded shape and first few rows.

```
# load all data
dataset = read_csv('household_power_consumption.txt', sep=';', header=0,
low_memory=False, infer_datetime_format=True, parse_dates={'datetime':[0,1]},
index_col=['datetime'])
# summarize
print(dataset.shape)
print(dataset.head())
```

Next, we can mark all missing values indicated with a '?' character with a NaN value, which is a float.

Data Preprocessing :




Preprocessing is a crucial step in data analysis and machine learning that involves cleaning, transforming, and organizing raw data to make it suitable for further analysis. It aims to enhance data quality and usability by removing noise, outliers, and inconsistencies. Preprocessing techniques often include data normalization, handling missing values, and feature scaling. This stage helps in preparing the data for tasks such as classification, regression, or clustering, ensuring more accurate and meaningful results. Proper data preprocessing can lead to improved model performance and insights in various fields, including healthcare, finance, and natural language processing.


Data preprocessing is an essential step in data analysis and machine learning, as it helps clean and format data for further analysis. In this example, I'll show you how to preprocess a CSV file with some basic operations. You'll need to have the pandas library installed to run this progra m .

```
import pandas as pd

# Load the CSV file into a DataFrame
df = pd.read_csv("data.csv")

# Handling missing values (NaN)
df.fillna(value=0, inplace=True) # Replace NaN with 0

# Drop rows with missing values
df. dropn a (in pla ce = Tru e )

# Removing leading/trailing spaces from text columns (Name)
df['Name'] = df['Name'].str.strip()

# Data type conversion (convert 'Age' and 'Salary' to integers)
df['Age'] = df['Age'].astype(int)
df['Salary'] = df['Salary'].astype(int)

# Save the preprocessed data to a new CSV file
df.to_csv("preprocessed_data.csv", index=False)
```

Conclusion :

"In conclusion, data preprocessing is an indispensable part of any data-driven project. It enhances the quality and reliability of the dataset, leading to more accurate and meaningful results in analysis and machine learning. Through techniques such as handling missing data, data cleaning, feature engineering, and data scaling, we can prepare the data for optimal model performance. Investing time and effort into data preprocessing pays off by improving the overall quality of insights and predictions, making it a critical phase in the data pipe lin e ."

**Thank You .**