INTRODUCTION:

Data preparation is a crucial and often time-consuming, one of the critical steps in the field of data analytics. This process involves cleaning, transforming, and organizing raw data into a format that is suitable for analysis. Effective data preparation is essential because the quality of your data directly impacts the quality and reliability of the insights or models and derive an example dataset called “public health_awareness.csv”

DATA LOADING:

- **Importing Libraries:**
  To begin the data analysis process, we first import essential libraries. In this case, we utilize the Pandas Library, a popular tool for data manipulation and analysis. The library is imported as ‘pd’ for convenient access to its functions and objects.

- **Loading the Dataset:**
  We load the dataset ‘public health_awareness.csv’ into a Pandas Data Frame. This step is essential as it brings the data in a structured format.

- **Initial Exploration:**
  A preliminary exploration of the dataset is conducted to get a overview and display the first few rows of the dataset to see its structure, column name

We also use the info() and describe() functions to gather valuable information about the dataset, including data types, summary statistics

DATA PREPROCESSING:

- **Handling Missing Values:**
  One of the most critical aspects of data preprocessing is addressing missing values. In our dataset, we employ a basic technique for handling missing data by removing rows with missing values.

- **Saving the Preprocessed Data:**
  Finally, the preprocessed dataset is saved as a new CSV file for future use. The file is named “Dataloading_preprocessed.csv”, and index = False argument ensures that the index column is not included in the file.

PYTHON SCRIPT:

```
import pandas as pd

data = pd.read_csv(“public health_awareness.csv”)

print(data.head())

print(data.info()) print (data.describe())
```

```
data = data.dropna()

data.to_csv("Dataloading  preprossed.csv". index=False)
```

CONCLUSION:

Data loading and preprocessing are fundamental steps in the data analysis process. It ensures that the data is in the right format, devoid of missing values, and ready for further exploitation, analysis and virtualization.