In today's digital world, data in various formats is valuable, as it is important for sources of information and insights. Raw data present in various forms, such as bytes, texts, multimedia, etc., is converted into many forms to be used by various organizations. Hence, cleaning and processing data is essential to derive meaningful information and insights from the data
.
Data wrangling is the process of converting raw and unprocessed data from one form to another to make it more recognizable and usable. It is known by many names, such as Data cleaning, munging, and remediation. Data wrangling is important for cleaning, structuring, and organizing data in a desired format for better business and useful information.

What is Data Wrangling?

Data Wrangling is known by many names, such as Data Cleaning, Data munging, and Data remediation. It is the process of collecting, cleaning, and converting raw data into a structured format for data analysis and decision-making process.
Hence, it is important to process these data and organize them to extract important pieces of information.
It also helps to increase the accuracy and readability of the raw data. With the help of data wrangling, more and more complex data can be handled easily once it is structured. Hence, data wrangling is important for all the big companies that heavily rely on data for their daily work. There are four significant steps in the data wrangling process.

Data Wrangling Process

Data wrangling takes a series of stages to process data in a desired format. Let us understand the complete process of Data wrangling below.

Data Collection: This is the first process where data is collected from various sources. There are many data sources, as it is present in various forms, such as electronic bytes, texts, audio, images, etc.

Data Cleaning: The data collected is generally in raw and unstructured format. In this stage, all the irregularities and inconsistencies in data are processed and removed.

Data Transformation: In this process, data is restructured into a structured format, which may involve converting data types, renaming, arranging data, etc.

Data Enrichment: In this stage, some additional information is fed into the dataset prepared.

Data Integration: Data, after processing from various sources, are combined into a single, unified dataset based on common factors.

Data Formatting: Data structures are now formatted into the form of tables, CSV files, or databases with the help of Excel, SQL, etc.

Data Publishing: This is the last stage of Data wrangling. It involves making data available to other users by giving them access to the application.

Benefits of Data Wrangling

As discussed above, data wrangling is important when dealing with large, unstructured datasets. Data wrangling has many benefits that we are going to discuss next.

1. Data Quality

Data wrangling helps us to improve the quality of our raw and unprocessed data by working on their errors, inconsistencies, and missing values and fixing them. This helps companies decode complex data easily and make good decisions in the interest of the company.

2. Consistency

As data wrangling structures our data in a usable format, it makes our data more consistent. It is very important for the business, as it helps in achieving the objectives and goals of the company. It is mostly used by companies that rely heavily on input from their users and process it.

3. Improved Efficiency

Implementing data wrangling improves the efficiency of the dataset as it gets easier to extract important information. Also, it reduces the work of data analysts by removing errors and inconsistencies in the dataset. They can easily focus on extracting useful insights.

4. Better Insights and Decision Making

As our data is well processed, extracting essential insights becomes easy. Also, decision-making becomes less time-consuming and productive as, in most cases, clean and processed data provides accurate data analysis.

5. Saves Time and Resources

Nowadays, there are many tools available that can highly automate the data-wrangling process and help reduce the time and resources used for the process. It not only saves time and effort but also reduces the cost by a significant amount.

Tools Used For Data Wrangling

There are many tools that can be used for performing various tasks of data wrangling. There are many programs that can automate the data cleaning process and validate the data during the process. Let us check out some of the important tools used in the process.

Data Wrangling Tools		
TOOL	**Description**	**Used For**
EXCEL POWER QUERY	It is a basic manual data wrangling tool used as an Excel feature.	Can be used for simple tasks.
OPENREFINE	This is an automated data-cleaning tool. It requires knowledge of programming.	Can be used for large-scale data-cleaning projects
TABULA	A versatile instrument capable of obtaining data from various types of documents, such as PDFs	Used for extracting data from documents.
GOOGLE DATAPREP	A Google data service that explores, cleans, and prepares data	Suitable for cloud-based data processing and cleaning.
DATA WRANGLER	A tool for cleaning and transforming data developed by Stanford University and used for data wrangling	Suitable for cleaning and changing various data sets.
TRIFACTA	A [data management](#) tool based on the cloud that offers intelligent data cleaning and transformation features	Ideal for large-scale and complex data wrangling tasks.
PANDAS	It provides data structures and functions to manipulate large datasets efficiently.	Suitable for programming-oriented data wrangling tasks
D3.JS	A JavaScript library for creating interactive data visualizations in web browsers	Suitable for visualizing and exploring cleaned data

Examples of Data Wrangling

There are many fields where data wrangling is used. Let us check some of the most common cases. \

- Deleting Unnecessary data
- Removing errors from the dataset
- Finding the missing fields in the dataset
- Merge data into one data set for data analysis
- Fixing inconsistencies