

Practical 2

AIM:

DataWranglingII

Create an “Academic performance” dataset of students and perform the following operations using Python.

1. Scan all variables for missing values and inconsistencies. If there are missing values and/or inconsistencies, use any of the suitable techniques to deal with them.
2. Scan all numeric variables for outliers. If there are outliers, use any of the suitable techniques to deal with them.
3. Apply data transformations on at least one of the variables. The purpose of this transformation should be one of the following reasons: to change the scale for better understanding of the variable, to convert a non-linear relation into a linear one, or to decrease the skewness and convert the distribution into a normal distribution.

THEORY:

Data Wrangling:

- Data wrangling is the process of **cleaning, organizing, and transforming raw data**
- Converts data into a **useful format for analysis and decision-making**
- Also known as **data cleaning or data munging**
- Helps handle **complex data efficiently**
- Methods vary based on **project goals and data type**

Benefits of Data Wrangling:

- Improves **data usability and quality**
 - Converts data into a **compatible format**
 - Helps in **quick data processing and automation**
 - Integrates data from **multiple sources** (databases, files, APIs)
 - Allows handling of **large volumes of data**
 - Makes data sharing easier
-

Data Wrangling Tools:

- **Excel / Power Query**
- **Tabula**
- **Google DataPrep**

CONCLUSION:

This experiment demonstrated data wrangling techniques such as handling missing values, detecting and treating outliers, and applying data transformation to improve the dataset for analysis.