**Data Cleaning**

**ROLL NO: 230701085**

**Objective:** To clean the diabetes dataset by handling missing values, outliers, and inconsistent data.

**Materials Needed:**

 Software Engineer Salaries dataset

 Python with libraries: pandas, numpy, matplotlib

**Procedure and Code:**

1. **Handling Missing Values:**

**CODE:**

import pandas as pd

import matplotlib.pyplot as plt

db=pd.read\_csv('Software Engineer Salaries.csv')

print(db.isnull().sum())

db.fillna(db.mean(),inplace=True)

print(db.isnull().sum())

**OUTPUT:**

Company 2

Company Score 0

Job Title 0

Location 13

Date 0

Salary 106

dtype: int64

**Handling Outliers:**

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

plt.figure(figsize=(10, 6))

sns.boxplot(data=db)

plt.show()

**OUTPUT:**

