43. Data Cleaning and Transformation

Task:

• a) Load the employee\_data.csv file into a Pandas DataFrame. The file contains the columns:

Employee ID, Full Name, Department, and Salary.

• b) Convert the Salary column to a numeric data type.

• c) Remove any rows where the Department column has missing values.

• d) Create a new column named First Name that extracts the first name from the Full Name column.

CODE:

import pandas as pd

employee\_df = pd.read\_csv("employee\_data.csv")

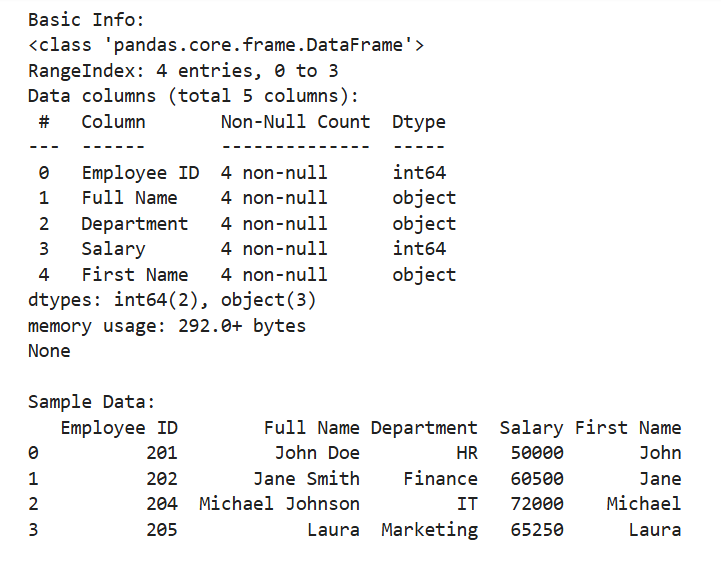
employee\_df['Salary'] = pd.to\_numeric(employee\_df['Salary'].replace('[\$,]', '', regex=True), errors='coerce')

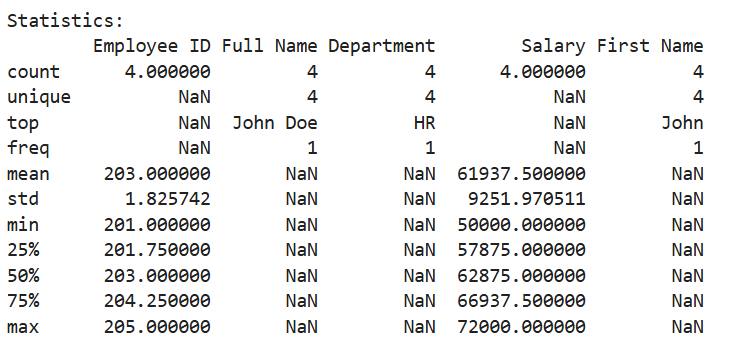
employee\_df = employee\_df.dropna(subset=["Department"])

employee\_df['First Name'] = employee\_df['Full Name'].str.split().str[0]

print(employee\_df.head())

OUTPUT:





Dataset:

|  |  |  |  |
| --- | --- | --- | --- |
| Employee ID | Full Name | Department | Salary |
| 201 | John Doe | HR | $50,000 |
| 202 | Jane Smith | Finance | $60,500 |
| 203 | Emily Davis | | $55,000 |
| 204 | Michael Johnson | IT | $72,000 |
| 205 | Laura | Marketing | $65,250 |
|  |  |  |  |