

By Muhamad Albatarni

AI Engineering Student | Python Automation Freelancer

Before & After Data Cleaning

Before Automation

ID	Name	Age	Date	Score
1	John Doe	25	01/15/2023	85
2	Jane Smith	20/02/2023	90	
3	John Doe	25	01-15-2023	85
4	Mike Brown	30	March 5 2023	88
5	22	10/03/2023		

(Note missing ages, inconsistent date formats, and duplicate rows)

After Automation

By *Muhamad Albatarni*

AI Engineering Student | Python Automation Freelancer

ID	Name	Age	Date	Score
1	John Doe	25	2023-01-15	85
2	Jane Smith	27	2023-02-20	90
4	Mike Brown	30	2023-03-05	88
5	22	2023-03-10	92	

(Duplicates removed, missing ages filled, dates standardized)

Sample Python Script

```
import pandas as pd

df = pd.read_csv('Messy.csv')
df_cleaned = df.drop_duplicates()
median_age = df_cleaned['Age'].median()
df_cleaned['Age'] = df_cleaned['Age'].fillna(median_age)
df_cleaned['Date'] = pd.to_datetime(df_cleaned['Date'],
errors='coerce', dayfirst=True)
df_cleaned['Date'] = df_cleaned['Date'].dt.strftime('%Y-%m-%d')
df_cleaned.to_csv('cleaned_data.csv', index=False)
```

By Muhamad Albatarni

AI Engineering Student | Python Automation Freelancer

Benefits of Automation

- Saves hours of manual Excel work
- Eliminates human error in data cleaning
- Standardizes data formatting for easy analysis
- Automates repetitive data processing tasks