```
import numpy as np
import pandas as pd

df=pd.read_csv("/content/drive/MyDrive/Student_data.csv")
df

before_rows = len(df)
print(before_rows)

missing_values_before = df.isnull().sum()
print(missing_values_before)

df['Region'] = df['Region'].fillna("Unknown").str.upper()

df['Marks'] = pd.to_numeric(df['Marks'], errors='coerce')
average_marks = df['Marks'].mean()
df['Marks'] = df['Marks'].fillna(average_marks)

df['Name'] = df['Name'].astype(str).str.strip()
df['Name'] = df['Name'].apply(lambda x: x.title() if x.strip()
else "Unnamed")

df['Date'] = pd.to_datetime(df['Date'],
errors='coerce').dt.strftime('%d-%m-%Y')
df.drop_duplicates(inplace=True)

df['Result'] = df['Marks'].apply(lambda x: 'Pass' if x >= 70 else
'Fail')

missing_values_after = df.isnull().sum()
print(missing_values_after)
```

```
after_rows = len(df)
print(after_rows)

df.to_csv("Student_data_cleaned.csv", index=False)

print("Summary Table\n")
for key, value in summary_table.items():
    print(f"{key}:\n{value}\n")
```

```
summary_table={
    'Rows Before': before_rows,
    'Rows After': after_rows,
    'Missing Before': missing_values_before,
    'Missing After': missing_values_after,
    'Pass Count': df[df['Result'] ==
'Pass'].groupby('Subject').size(),
    'Fail Count': df[df['Result'] ==
'Fail'].groupby('Subject').size()
}
summary_table
```

```
from google.colab import files
files.download("Student_data_cleaned.csv")
```