Ex.No :	Visualize the data with appropriate charts and perform a simple linear regression, interpreting the results using a R Program.
Date :	

AIM

To import a dirty dataset into Python and perform data cleansing: handle missing values and correct data types.

ALGORITHM

Step 1: Start the Python program.

Step 2: Import pandas and numpy.

Step 3: Create or read the dirty dataset.

Step 4: Inspect with head() and info().

Step 5: Fill or remove missing values (use mean/imputation or drop).

Step 6: Convert columns to correct types (astype).

Step 7: Verify and print cleaned dataset.

Step 8: Stop the program.

PROGRAM

```
import pandas as pd
import numpy as np
data = {
   'Name': ['sam', 'swethae', 'niki', 'ram', 'vino'],
   'Age': ['21', '22', np.nan, '20', '23'],
   'Marks': [85, np.nan, 90, 75, '88']
}
df = pd.DataFrame(data)
print("---- Original Dataset ----")
print(df)
print("\nData Info before cleaning:")
print(df.info())
df['Age'].fillna(df['Age'].astype(float).mean(), inplace=True)
df['Marks'].fillna(df['Marks'].astype(float).mean(), inplace=True)
df['Age'] = df['Age'].astype(float)
df['Marks'] = df['Marks'].astype(float)
print("\n---- Cleaned Dataset ----")
print(df)
```

```
print("\nData Info after cleaning:")
print(df.info())
```

OUTPUT

```
---- Original Dataset ----
      Name  Age Marks
0      sam   21    85
1  swethae   22   NaN
2     niki  NaN    90
3      ram   20    75
4     vino   23    88
```

```
---- Cleaned Dataset ----
      Name   Age  Marks
0      sam  21.0   85.0
1  swethae  22.0   84.5
2     niki  21.5   90.0
3      ram  20.0   75.0
4     vino  23.0   88.0
```

RUBRICS

	Max Marks	Marks Obtained
Understanding of concept	10	
Code Quality	40	
Output	20	
Ethical consideration & Societal Impact	10	
Vice-Voce	10	
Report Quality	10	
Total	100	

RESULT

Thus missing ages and marks were replaced by their respective column means.The Age and Marks columns were successfully converted to numerical types.The dataset was cleaned and made ready for analysis