Below is a Python example to Clean data for leads' enquiry base and traffic sources using a CSV file. The cleaning process typically involves steps such as:

1. Loading the data into a DataFrame.
2. Identifying and handling missing data (e.g., filling, dropping, or replacing missing values).
3. Correcting data types.
4. Normalizing or standardizing categorical values (e.g., traffic sources).
5. Removing duplicates.
6. Handling any outliers or errors in data (depending on context).

Let's assume the CSV file has the following columns:

- `Lead_ID`: Unique identifier for each lead.
- `Enquiry_Date`: Date when the enquiry was made.
- `Traffic_Source`: The source through which the lead was acquired (e.g., "Google", "Facebook", "Organic").
- `Lead_Status`: Whether the lead was converted or not.
- `Lead_Amount`: Value associated with the lead (e.g., revenue or potential value).

Code Example :-

```
import pandas as pd

# Load the CSV file
df = pd.read_csv('leaddata.csv')

# Show initial data for review
print("Initial Data")
print(df.head())

# Step 1: Handling missing data
# Fill missing Lead_Amount with 0
df['Lead_Amount'] = df['Lead_Amount'].fillna(0)

# Drop rows where 'Traffic_Source' or 'Lead_ID' is missing
df = df.dropna(subset=['Traffic_Source', 'Lead_ID'])

# Step 2: Standardize traffic sources (convert to lowercase and strip extra
spaces)
df['Traffic_Source'] = df['Traffic_Source'].str.lower().str.strip()

# Step 3: Correcting data types
# Ensure that 'Enquiry_Date' is a datetime type
```

```
df['Enquiry_Date'] = pd.to_datetime(df['Enquiry_Date'], errors='coerce')

# Ensure 'Lead_Amount' is numeric
df['Lead_Amount'] = pd.to_numeric(df['Lead_Amount'], errors='coerce')

# Step 4: Removing duplicates based on 'Lead_ID'
df = df.drop_duplicates(subset=['Lead_ID'])

# Step 5: Handling any outliers in 'Lead_Amount' (example: if amount > 1,000,000
is considered an outlier)
df = df[df['Lead_Amount'] <= 1000000]

# Step 6: Convert 'Lead_Status' to a consistent format (e.g., 'converted' or 'not
converted')
df['Lead_Status'] = df['Lead_Status'].str.lower().str.strip()

# Final cleaned data
print("Cleaned Data")
print(df.head(10))

# Save cleaned data to a new CSV file
df.to_csv('cleaned_leads_data.csv', index=False)
```

output:-

Initial Data

	Lead_ID	Enquiry_Date	Traffic_Source	Lead_Status	Lead_Amount
0	1	2024-12-20	Google	Converted	1500
1	2	2024-12-21	Facebook	Not Converted	200
2	3	2024-12-22	google	converted	500
3	4	2024-12-23	Instagram	Not Converted	300
4	5	2024-12-24	NaN	Converted	1200

Cleaned Data

	Lead_ID	Enquiry_Date	Traffic_Source	Lead_Status	Lead_Amount
0	1	2024-12-20	google	converted	1500
1	2	2024-12-21	facebook	not converted	200
2	3	2024-12-22	google	converted	500
3	4	2024-12-23	instagram	not converted	300
5	7	2024-12-26	google	converted	350
7	8	2024-12-27	organic	not converted	600
8	9	NaT	google	converted	1000
9	10	2024-12-29	facebook	converted	2500
10	11	2024-12-30	facebook	NaN	500
11	12	2024-12-31	organic	converted	800