**Preprocessing Steps for Global Inflation Dataset**

**1️⃣ Load and Inspect Data**

* **Tool Used:** pandas.read\_excel()
* **File Format:** Excel .xlsx
* **Initial Checks:**
  + Reviewed column types (country, indicator, years)
  + Verified presence of year-wise inflation data (1980–2024)

**2️⃣ Rename and Standardize Columns**

* Renamed columns like Country Name to country\_name (if needed).
* Standardized column names to lowercase with underscores for consistency.

**3️⃣ Filter Relevant Time Period**

* Focused on **last 10 years**: 2015–2024

python

CopyEdit

latest\_year = max([col for col in df.columns if isinstance(col, int)])

years\_of\_interest = list(range(latest\_year - 9, latest\_year + 1))

df\_filtered = df[['country\_name', 'indicator\_name'] + years\_of\_interest]

**4️⃣ Handle Missing Values**

* Identified missing values using .isnull().sum()
* Observations:
  + Higher missing values in early years (1980s–1990s)
  + Minimal missing values in 2015–2024 range
* Action:
  + No imputation done for analysis (kept nulls to preserve accuracy)
  + Suggested imputation for modeling (e.g., forward fill, regional average)

**5️⃣ Remove Duplicates**

* Checked and confirmed:

python

CopyEdit

df.duplicated().sum() # 0

**6️⃣ Convert Columns**

* Ensured all year columns are **numeric (float)**:

python

CopyEdit

df[years\_of\_interest] = df[years\_of\_interest].apply(pd.to\_numeric, errors='coerce')

**7️⃣ Create Derived Metrics**

* Created **Inflation Volatility** metric:

python

CopyEdit

df['inflation\_volatility'] = df[years\_of\_interest].std(axis=1)

* Useful for sorting countries by stability/instability.

**8️⃣ Reshape for Visualization**

* Converted wide to long format using melt():

python

CopyEdit

df\_melted = df.melt(id\_vars=['country\_name'], value\_vars=years\_of\_interest,

var\_name='year', value\_name='inflation')

* Enables line plots in **Seaborn**, **Matplotlib**, or **Power BI**

**✅ Summary of Key Actions**

| **Step** | **Action** |
| --- | --- |
| ✅ Load Data | Loaded Excel using pandas |
| ✅ Rename Columns | Standardized naming |
| ✅ Time Filter | Selected 2015–2024 |
| ✅ Missing Handling | Retained or flagged |
| ✅ Duplicates | None found |
| ✅ Type Casting | Converted years to numeric |
| ✅ Derived Metric | Std Dev of inflation |
| ✅ Long Format | Used for plots |