**Data Cleaning Report: *Global Inflation Dataset***

**Basic Structure**

* **Total Rows:** 196
* **Total Columns:** 47
* **Key Columns:** country\_name, indicator\_name, 1980–2024

**Missing Values Summary**

| **Column** | **Missing Values** | **Missing (%)** |
| --- | --- | --- |
| 1980 | 56 | 28.57% |
| 1981 | 52 | 26.53% |
| 1982 | 51 | 26.02% |
| 1983 | 51 | 26.02% |
| 1990 | 46 | 23.47% |
| ... | ... | ... |
| 2023 | 4 | 2.04% |
| 2024 | 5 | 2.55% |

**Observation**: Data from earlier years (1980–1992) has high missingness (20–30%). Recent years (2010–2024) have minimal missing values (<3%).

**Duplicate Records**

* **Duplicates Found:** **0**

No duplicate rows detected.

**Summary Statistics (Recent Years Only)**

| **Year** | **Mean Inflation (%)** | **Std Dev** | **Max (%)** | **Min (%)** |
| --- | --- | --- | --- | --- |
| 2020 | 19.83 | 173.72 | 2355.10 | -2.60 |
| 2021 | 16.58 | 117.15 | 1588.50 | -3.00 |
| 2022 | 13.62 | 25.28 | 193.40 | -3.20 |
| 2023 | 13.74 | 39.67 | 360.00 | -0.80 |
| 2024 | 9.31 | 25.20 | 222.40 | 1.20 |

**Note**: Extremely high values in max inflation suggest **outliers** or countries experiencing **hyperinflation**.

**Recommendations**

1. **Drop or Impute Early Years (1980–1992)**: Due to high missing values.
2. **Outlier Treatment**: Consider winsorizing or log-scaling for visualization.
3. **Consistency Checks**:
   * Ensure country\_name values are standardized.
   * Check inflation values that are negative or suspiciously large.

Would you like this as a downloadable **Excel or PDF file** for inclusion in your GitHub or Power BI project folder?

**You said:**

in a pdf