**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 02 Oct 2025 |
| Name | Muhammad Yasir |
| Team ID | xxxxxx |
| Project Title | Power BI Inflation Analysis: Journeying Through Global Economic Terrain |
| Maximum Marks | 3 Marks |

| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| --- | --- | --- | --- |
| Dataset | Missing inflation rate values for some years (e.g., Albania, Syria) | High | Impute missing values with median per country, or regional averages where data gaps are large. |
| Dataset | Duplicate country-year records | Moderate | Remove duplicates by keeping the latest or most reliable record based on source. |
| Dataset | Inconsistent country names (e.g., “CÃ´te d'Ivoire” instead of “Côte d'Ivoire”) | Moderate | Standardize names using ISO country codes or mapping table. |
| Dataset | Year columns stored in wide format (1980–2024) | Low | Unpivot year columns into long format (Country, Year, Inflation Rate). |
| Dataset | Inflation values stored as text in some cells | Moderate | Convert text fields to numeric (decimal/percentage). |
| Dataset | Extreme outliers (e.g., Venezuela hyperinflation) | Low/Moderate | Keep outliers but flag them; allow filtering in dashboard to prevent distortion. |
| Dataset | Blank or inconsistent region assignments | Moderate | Create DimCountry with Region & Subregion mapping, join via lookup. |
| Dataset | Different decimal formats (comma vs dot) | Low | Normalize all inflation values to a consistent decimal separator. |

## Severity Scale

* **High:** Prevents accurate analysis, must be resolved immediately.
* **Moderate:** Affects some analyses, should be fixed or mitigated.
* **Low:** Minor issue, can be documented or flagged without urgent correction.

✅ **Outcome:** This Data Quality Report helps track issues systematically, ensures transparency in cleaning, and provides clear action plans for resolution.