**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 02 Oct 2025 |
| Team ID | xxxxxx |
| Project Title | Power BI Inflation Analysis: Journeying Through Global Economic Terrain |
| Name | Muhammad Yasir |
| Maximum Marks | 10 Marks |

## Data Overview

* **Data Source:** CSV, Excel
* **Dataset Size: 196,** 48
* **Key Columns:** Country Name, Indicator Name ,Years, Inflation Rate.
* **Time Range:** 1980–2024
* **Initial Observations:** wide format with years as columns, contains missing values, has regional/country mapping needed

## Data Cleaning

* **Missing Values:**
  + Identify % of missing data per column.
  + Strategy: fill with median (per country), fill with regional average .
* **Duplicates:**
  + Detect duplicate rows based on keys (Country, Year).
  + Remove or resolve as per requirement.
* **Error Corrections:**
  + Fix spelling issues in country names.
  + Standardize region labels e.g. “U.S.” vs “United States”.

## Data Transformation

* **Reshape Data:**
  + Unpivot year columns (1980–2024) into a long format with columns: Country, Year, Inflation Rate.
* **Filtering and Sorting:**
  + Remove irrelevant columns e.g. metadata.
  + Sort data chronologically for time series.
* **Calculated Columns:**
  + Median Inflation per country.
  + YoY % Change.
  + Decade grouping (e.g., 1980s, 1990s).
  + Region and Sub region mapping.

## Data Type Conversion

* Convert Year → Integer or Date.
* Convert Inflation Rate → Decimal/Percentage.
* Ensure categorical fields Country, Region are stored as Text.

## Column Splitting and Merging

* **Splitting:**
  + Separate composite fields (e.g., “Country, Region” into two columns).
* **Merging:**
  + Combine Year + Quarter into a single Date field (if needed).
  + Concatenate Country + Year for unique identifiers.

## Data Modeling

* **Fact Table:** Inflation data (Country, Year, Inflation Rate, Median of inflation rate).
* **Dimension Tables:**
  + Dim Country : Country, Region, Sub region.
  + Dim Time : Year, Decade, Half-Decade.
* **Relationships:** Link Fact Inflation with Dim Country and Dim Time.
* **Measures:** Average Inflation, Median Inflation, Max/Min, YoY Change.
* **Schema:** make a Star schema.

## Save Processed Data

* **Export Formats:** Clean Excel file, CSV, or Power BI Dataflow.
* **Versioning:** Keep track of processing date/version.
* **Storage:** Save to shared folder or cloud repository.

**Documentation & Validation**

* Document all transformations in Power Query steps.
* Validate by comparing summary statistics (mean, median, min, max) before vs after cleaning.
* Ensure consistency between regions, countries, and years.

**Outcome:** A reliable, cleaned, and structured dataset ready for visualization, reporting, and advanced analysis.