Data Cleaning Process – Gen-Z Career Aspirations Dataset

1. Eliminate Null Values

- Identify missing values using Excel filters or ISBLANK() formula.

- Fill missing numerical values with 0 or a relevant calculated value.

- For categorical data, replace with "Not Specified" or another placeholder.

- Delete rows where most of the important fields are blank.

---

2. Use the Given Gen-Z Dataset

- Work only on the provided **"Your Career Aspirations (Responses)"** dataset.

- Keep the original file untouched in the raw_data/ folder for reference.

---

3. Splitting into Rows

- For cells containing multiple comma-separated responses:

    o Use Excel's **Text to Columns** (if you want to split into columns first).

    o Or use **Power Query → Split Column → By Delimiter → Split into Rows**.

---

4. Data Formatting

- Convert date/time fields to proper **DateTime** format (YYYY-MM-DD HH:MM:SS).

- Ensure numeric columns are stored as numbers (no text format).

- Apply consistent number formatting for percentages, currency, etc.

---

5. Field Items Standardization (Data)

- Remove extra spaces using TRIM() in Excel.

- Ensure consistent capitalization (e.g., Title Case for text).

- Replace inconsistent values with standardized ones.

    o Example: "india", "India ", "INDIA" → "India"

---

6. Field Headers Standardization (Column)

- Remove extra spaces in column names.

- Convert all column names to lowercase.

- Replace spaces with underscores (_).

- Remove special characters (?, -, .).

  - Example: "Your Current Country." → your_current_country

🛠 Skills Applied

- **Microsoft Excel**

- **Excel Formulas** (TRIM(), ISBLANK(), TEXT())

- **Data Standardization and Cleaning in Excel**

- **Data Cleaning Techniques**

- **Data Transformation (Splitting into Rows)**

- **Field Header and Field Item Standardization**