**Documentation on Data Cleaning, Transformation, and Workflow in Power BI**

**1. Introduction**

Power BI is a Business Intelligence tool by Microsoft that enables data connection, cleaning, transformation, modelling, visualization, and reporting.  
The **ETL (Extract, Transform, Load)** process in Power BI is primarily performed in **Power Query Editor** before data is loaded into the data model. Proper **documentation** ensures data accuracy, reproducibility, and transparency.

**2. Data Cleaning in Power BI**

Data cleaning is the process of removing inconsistencies, handling missing values, and preparing raw data for analysis. This is done in **Power Query Editor**.

**🔹 Common Data Cleaning Tasks:**

1. **Remove Duplicates**
   * Use **Remove Duplicates** option in Power Query to ensure unique records.
2. **Handle Missing/Null Values**
   * Replace null values with defaults (e.g., 0, “Unknown”) using **Replace Values**.
   * Remove rows/columns with too many missing values.
3. **Trim and Clean Text**
   * Use **Trim** to remove leading/trailing spaces.
   * Use **Clean** to remove non-printable characters.
4. **Change Data Types**
   * Ensure proper types: dates → Date type, numbers → Decimal/Whole, categorical → Text.
5. **Split/Extract Columns**
   * Use **Split Column by Delimiter** for structured text (e.g., Full Name → First Name + Last Name).
6. **Filter Out Unwanted Data**
   * Remove unnecessary rows, columns, or apply conditions.
7. **Standardize Data**
   * Convert all text to lowercase/uppercase.
   * Format dates consistently (e.g., YYYY-MM-DD).

**3. Data Transformation in Power BI**

Transformation reshapes data into a usable structure for analysis. It ensures data models are efficient, consistent, and ready for visualization.

**🔹 Key Transformation Steps:**

1. **Merge Queries (Joins)**
   * Combine tables using relationships: Inner, Left, Right, Full Outer.
2. **Append Queries**
   * Stack multiple datasets with similar structure (e.g., monthly sales data).
3. **Pivot & Unpivot Data**
   * Pivot: Turn rows into columns.
   * Unpivot: Turn columns into attribute-value pairs.
4. **Grouping & Aggregation**
   * Summarize data by categories (e.g., sales by region).
5. **Calculated Columns & Measures (DAX)**
   * Create new fields using DAX formulas (e.g., Profit = Sales – Cost).
6. **Hierarchy Creation**
   * Build drill-down structures (e.g., Year → Quarter → Month → Day).
7. **Data Normalization/Denormalization**
   * Transform data into star schema for efficient querying.
8. **Apply Business Rules**
   * Example: Categorize sales as “High”, “Medium”, “Low” based on thresholds.

**4. Documentation Workflow in Power BI**

Documentation ensures others can understand, replicate, and maintain your Power BI project.

**🔹 Recommended Workflow:**

1. **Requirement Gathering**
   * Define project scope, data sources, KPIs, business questions.
2. **Data Source Documentation**
   * List all sources (Excel, SQL, APIs, CSVs).
   * Mention connection method (Direct Query / Import / Live).
3. **ETL Process Documentation**
   * Document each **cleaning and transformation step** in Power Query.
   * Use **“Applied Steps” pane** as a built-in tracker.
   * Export M-code (Power Query script) for technical documentation.
4. **Data Model Documentation**
   * Show table relationships (ER diagram).
   * Describe calculated columns, measures, and hierarchies.
5. **Version Control**
   * Maintain different versions of reports for audit purposes.
6. **Report Documentation**
   * Provide explanation for each dashboard (what KPI, filters, and slicers mean).
7. **Performance Optimization**
   * Document techniques like removing unused columns, aggregations, incremental refresh.
8. **Final User Guide**
   * Create a README for end-users on how to use dashboards and interpret data.

**5. Best Practices**

* Always use **meaningful column names** after cleaning.
* Use **comments in DAX** for calculated measures.
* Keep **transformation steps minimal** to avoid performance issues.
* Maintain a **data dictionary** (field names, descriptions, types).
* Use **separate queries for staging vs. final tables** for clarity.
* Validate transformed data with summary statistics.

1. **Clean**: Remove duplicates, fix date formats, replace null values.
2. **Transform**: Merge Sales with Customers & Products, create profit column, unpivot monthly sales.
3. **Model**: Build star schema (Fact Sales linked to Dim Customer, Dim Product).
4. **Visualize**: Create dashboards for Revenue, Profit, Customer Segmentation.
5. **Document**: Maintain an ETL log, data dictionary, and report usage guide.

**7. Tools Supporting Documentation**

* **Power Query (M Code Export)** – for transformation scripts.
* **Data Model View** – for relationship diagrams.
* **External Tools** (like DAX Studio, Tabular Editor) – for advanced documentation.
* **Excel / Word Report** – maintain external documentation if required.