1. Identify Cleaning Tasks

Below are the cleaning tasks you can perform in Power BI's Power Query Editor, organized by column:

General Data Cleaning Tasks

- **Remove Duplicates**: Ensure Employee ID has unique values.

- **Trim Whitespace**: Remove extra spaces in text fields like Full Name, Department, and Designation.

- **Ensure Consistent Data Types**: Set appropriate data types (e.g., text, number, date).

- **Remove Nulls or Missing Values**: Replace or fill missing entries in critical fields.

Column-Specific Cleaning Tasks

1. **Employee ID**:
   - Ensure there are no duplicate IDs.
   - Replace any null values with a default identifier like "Unknown".

2. **Full Name**:
   - Capitalize the first letter of each name (if not already).
   - Remove any stray numbers or special characters that don't belong.

3. **Department**:
   - Replace typos or misclassifications with consistent categories.
     - e.g., Correct mismatches like "Humon Resources" to "Human Resources".
   - Filter out blank entries and replace them with "Not Assigned".

4. **Designation**:
   - Standardize capitalization.
   - Replace missing values with "Unassigned" or "Undefined".

5. **Date of Joining**:

- Identify invalid dates (e.g., 1900 or before) and replace them with "01-01-2000" or other logical values.
- Ensure all dates are in YYYY-MM-DD format.

6. **Annual Salary**:
- Replace nulls with a department average using Power Query's grouping and aggregation.
- Ensure all values are numeric with appropriate formatting (e.g., $123,456.78).

7. **Performance Score**:
- Replace nulls with the average score.
- Remove outliers outside the range (e.g., scores beyond 1–5).

8. **Leaves Taken**:
- Replace nulls with 0, assuming employees did not take leave.
- Ensure values do not exceed the total available (e.g., more than 24).

9. **Gender**:
- Fix mismatches ("M" → "Male", "F" → "Female").
- Replace blank fields with "Not Disclosed".

10. **Contact Number**:
- Replace invalid numbers with "9999999999".
- Remove any country codes or special characters (e.g., +1, -).

11. **Email**:
- Identify and replace invalid or blank emails with "noemail@domain.com".

12. **Status**:
- Ensure only allowed values ("Active", "Terminated", "Resigned").
- Replace typos or blank statuses with "Unknown".

13. **Termination Date**:
- Replace dates for "Active" employees with null.

14. **Work Location**:
- Standardize location names (e.g., "NewYork" → "New York").

- o   Replace nulls with "Remote" as the default.

15. **Employment Type**:

- o   Replace any unknown types with "Full-Time" by default.

- o   Validate all entries fall into expected types: ["Full-Time", "Part-Time", "Contract", "Internship"].

16. **Education Level**:

- o   Replace blank or invalid entries with "Not Provided".

- o   Correct inconsistent abbreviations (e.g., "HS" → "High School").

17. **Bonus/Benefits**:

- o   Replace nulls with department averages.

- o   Ensure values are numeric.

18. **Supervisor ID**:

- o   Replace nulls with "No Supervisor Assigned".

- o   Validate that supervisor IDs match existing employee IDs where applicable.

19. **Training Hours**:

- o   Replace nulls with 0.

- o   Cap values to a reasonable limit (e.g., 100 hours).

20. **Team Size**, **Project Count**:

- o   Replace blanks with department or role-based averages.

21. **Marital Status**:

- o   Standardize entries (e.g., "Sgl" → "Single").

- o   Replace nulls with "Unknown".

22. **Health Insurance**:

- o   Replace nulls with "Not Provided".

23. **Retirement Contributions**:

- o   Replace missing values with the mode (most frequent value).

24. **Nationality**:

- o   Standardize values (e.g., "AMERICAN" → "American").

- o Replace blanks with "Not Specified".

25. **Recent Promotion**:

- o Replace nulls with "No".