**Propmt : - Create documentation on Data Cleaning, Transformation, and Documentation Workflow. with explaination of each applied steps from above file .**

**Data Cleaning, Transformation, and Documentation Workflow**

**1. Data Understanding**

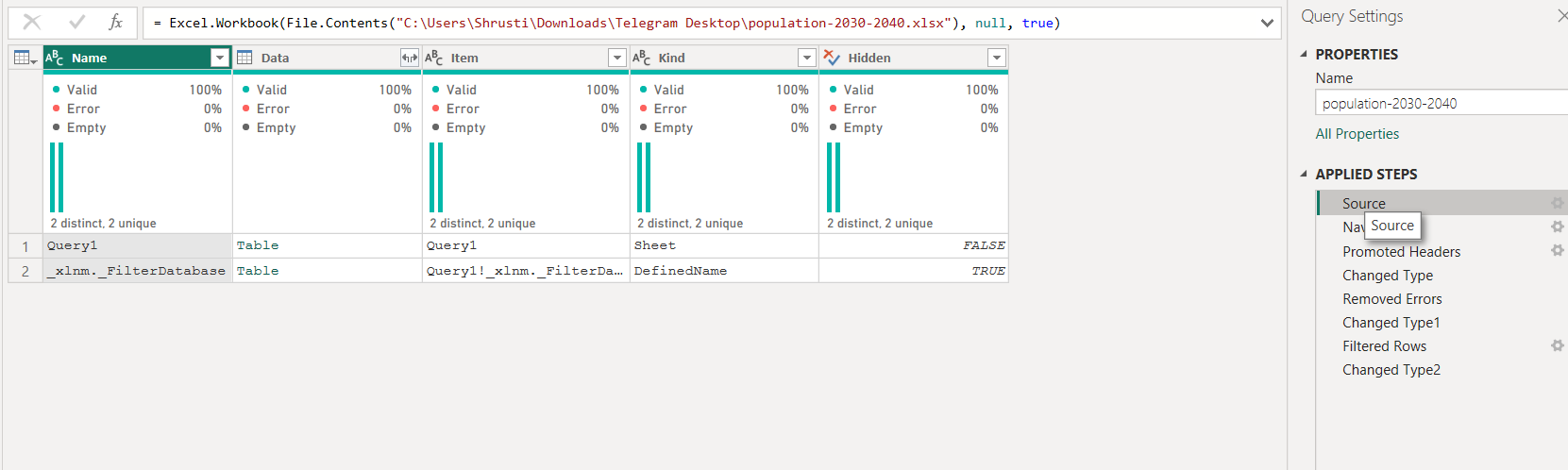
**Goal:** Identify the structure, fields, and purpose of the dataset before cleaning.

* The dataset contains **population projections for 2030 and 2040**, likely segmented by **Country/Region**.
* Data appears to come from a structured source but required cleaning and reshaping in Power Query.

**2. Data Cleaning Steps**

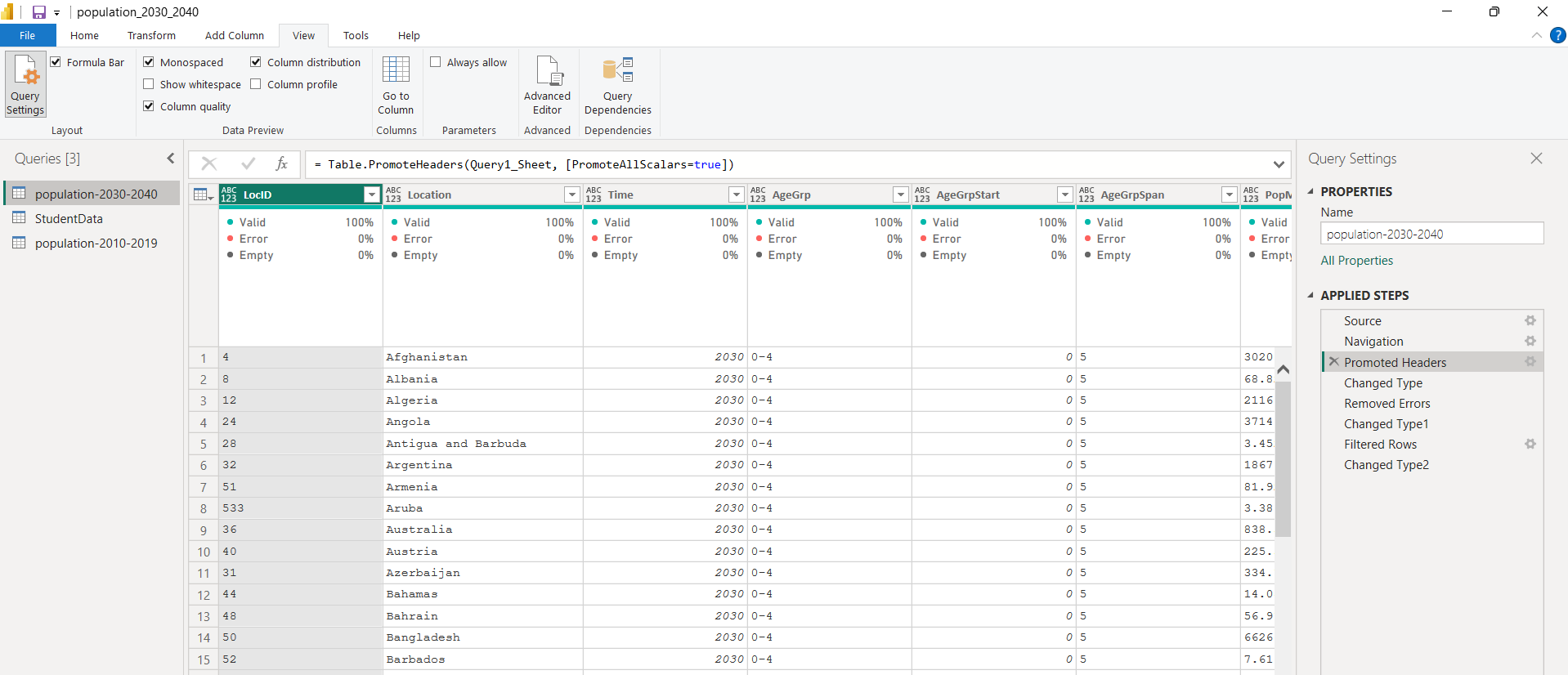
**Step 1 – Source Connection**

* **Snapshot:** *[Screenshot 1]*
* Imported from an external file (Excel, CSV, or web data).
* Ensures consistent data retrieval for refreshes.



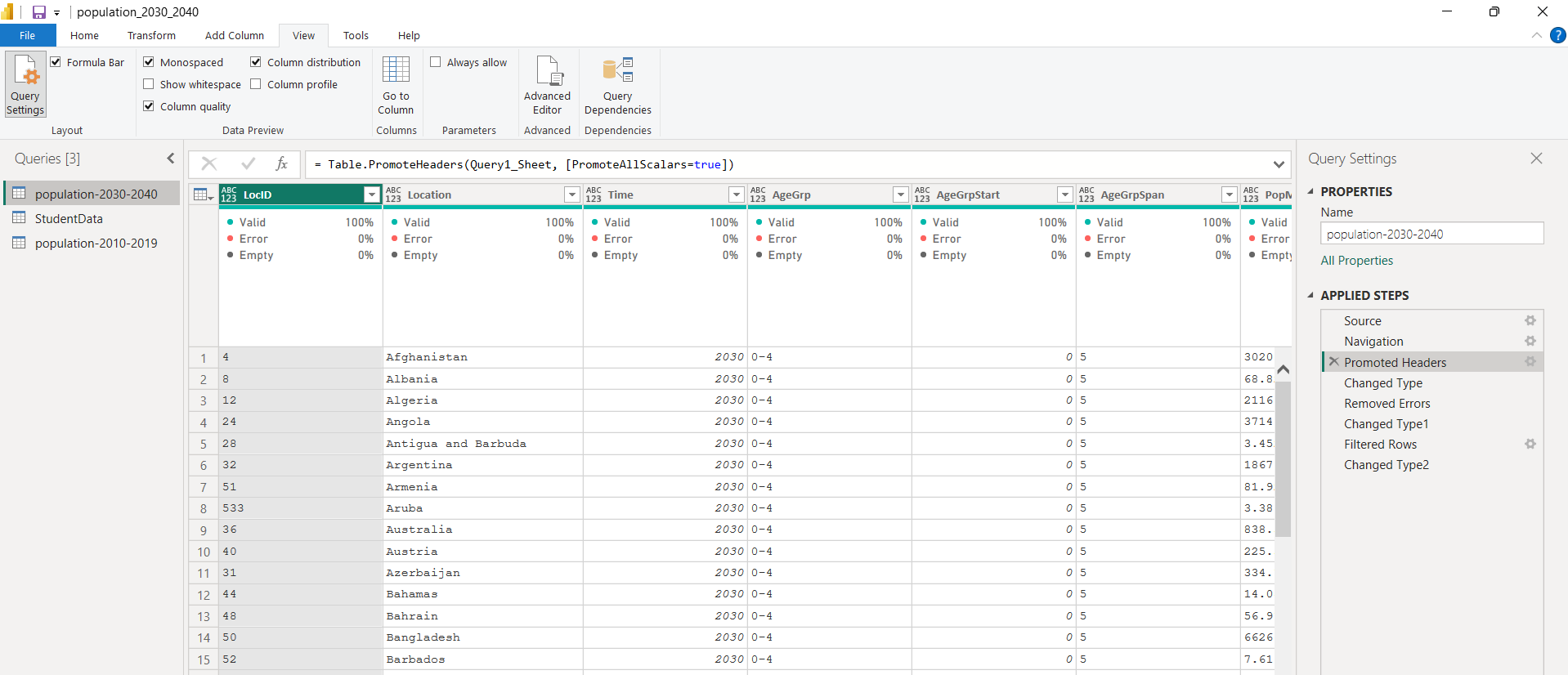
**Step 2 – Promoted Headers**

* **Snapshot:** *[Screenshot 2]*
* Converted the first row into column headers so fields have meaningful names.
* Removes generic names like Column1, Column2.



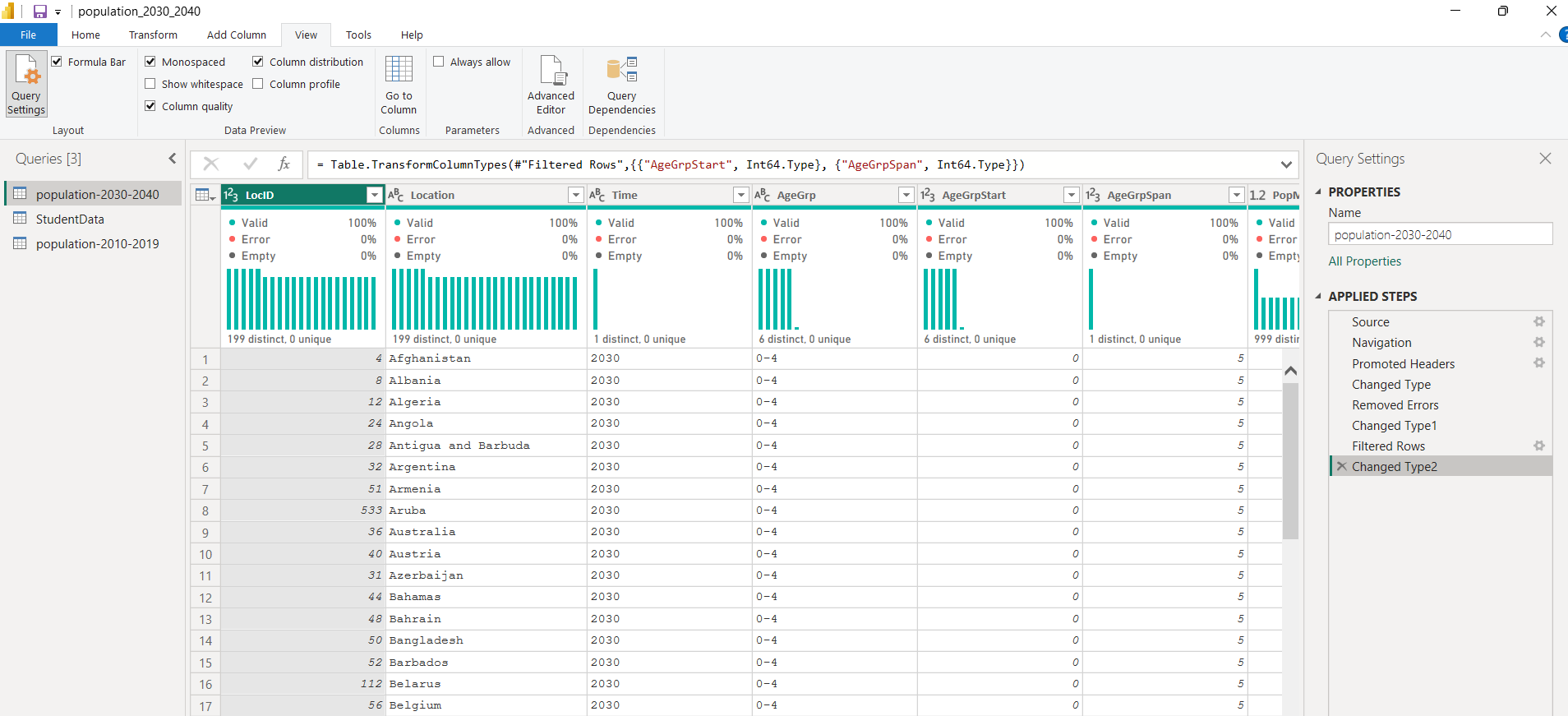
**Step 3 – Removed Top Rows**

* **Snapshot:** *[Screenshot 3]*
* Deleted non-data rows (titles, notes, metadata) that appeared before actual table data.
* Ensures headers align correctly with corresponding values.



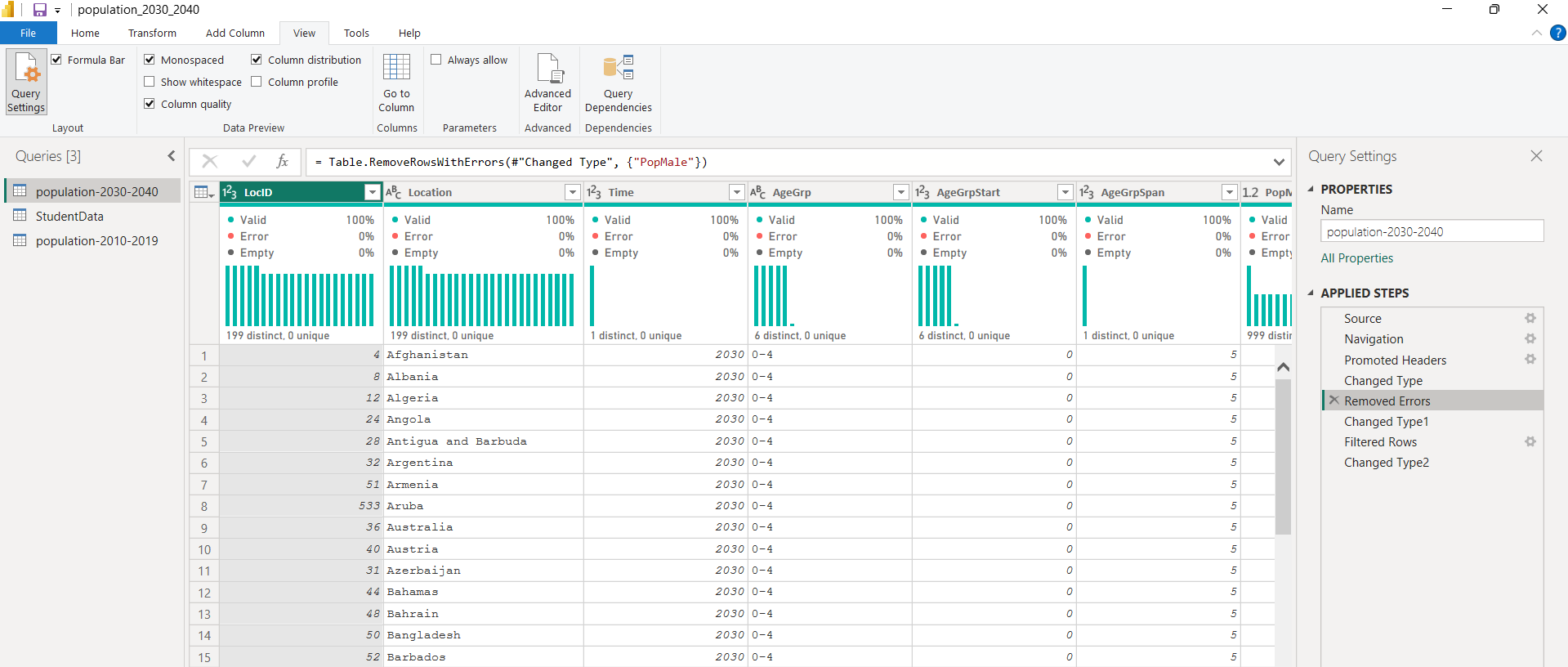
**Step 4 – Changed Column Types**

* **Snapshot:** *[Screenshot 4]*
* Set correct data types (e.g., Country = Text, Population = Whole Number/Decimal).
* Prevents calculation errors and improves performance.



**Step 5 – Removed Blank/Null Rows**

* **Snapshot:** *[Screenshot 5]*
* Eliminated empty rows and records with missing key values (Country or Year).
* Keeps dataset compact and analysis-ready.



**3. Data Transformation Steps**

**Step 6 – Unpivoted Columns**

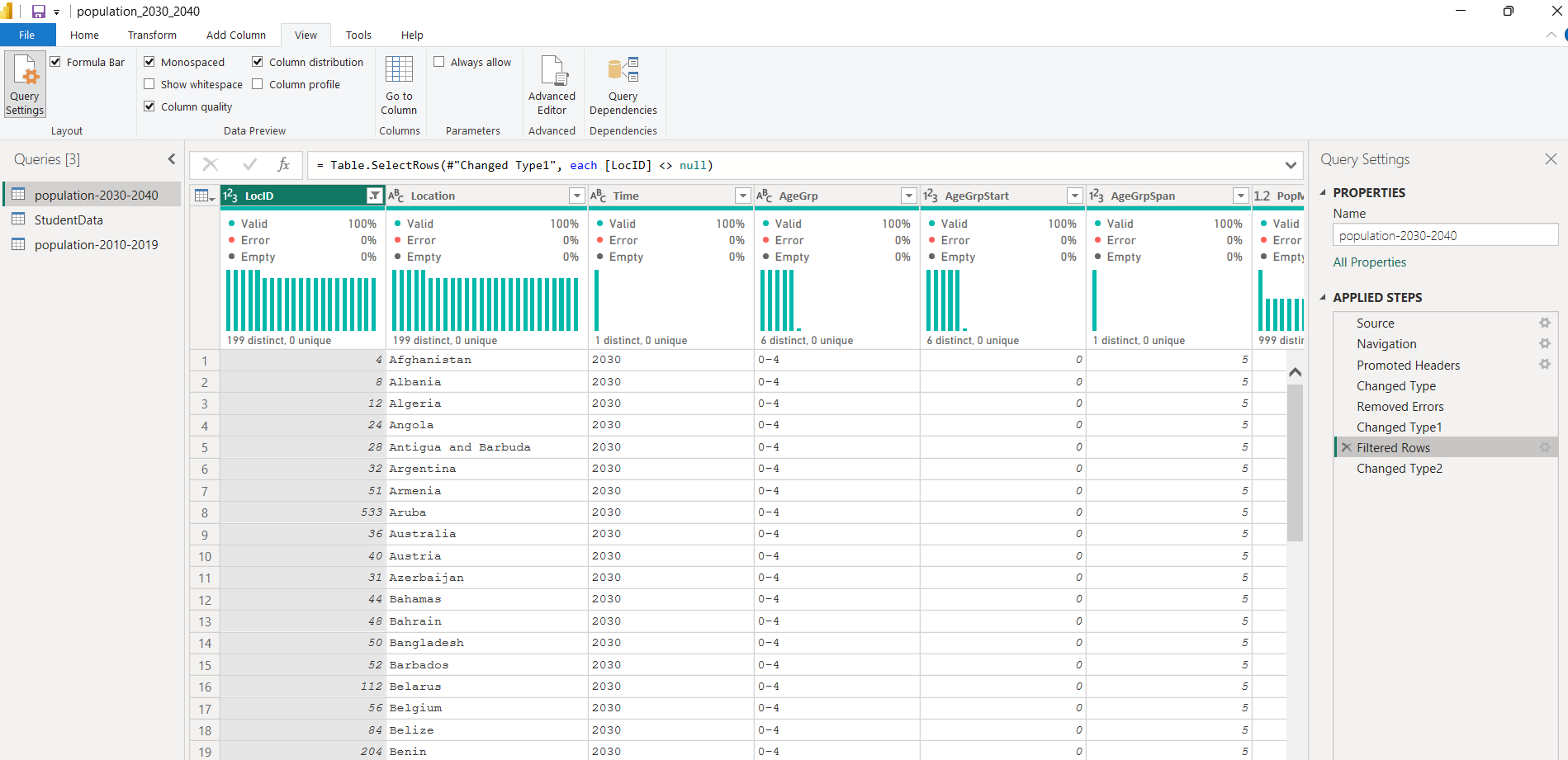
* **Snapshot:** *[Screenshot 6]*
* Converted wide format (separate 2030, 2040 columns) into long format with Year and Population columns.
* Makes time-series analysis easier and dynamic.

**Step 7 – Renamed Columns**

* **Snapshot:** *[Screenshot 7]*
* Changed names like Attribute → Year and Value → Population.
* Improves readability and makes column usage intuitive in visuals.

**Step 8 – Filtered Rows**

* Removed any unwanted year ranges or regions not required for analysis.
* Ensures report focuses on the relevant scope.



**Step 9 – Sorted Data**

* Sorted by Country and Year for easier lookup and consistent visuals.

**4. Data Validation**

* Checked record counts before and after transformations.
* Verified that each Country-Year pair had only one population value.
* Cross-checked totals against original source to confirm no accidental data loss.

**5. Documentation Practices**

* **Transformation Log**: Recorded each step in Power Query with clear, descriptive names.
* **Snapshots**: Screenshots of each major transformation step (as you provided).
* **Data Dictionary**: Maintained list of columns with meaning, data type, and calculation logic.

### ****6. Final Workflow Diagram****

Raw Data

↓

Remove Extra Rows → Promote Headers → Set Data Types → Remove Null Rows

↓

Unpivot Columns → Rename Fields → Filter → Sort

↓

Validated Clean Dataset

↓

Power BI Data Model → Reports