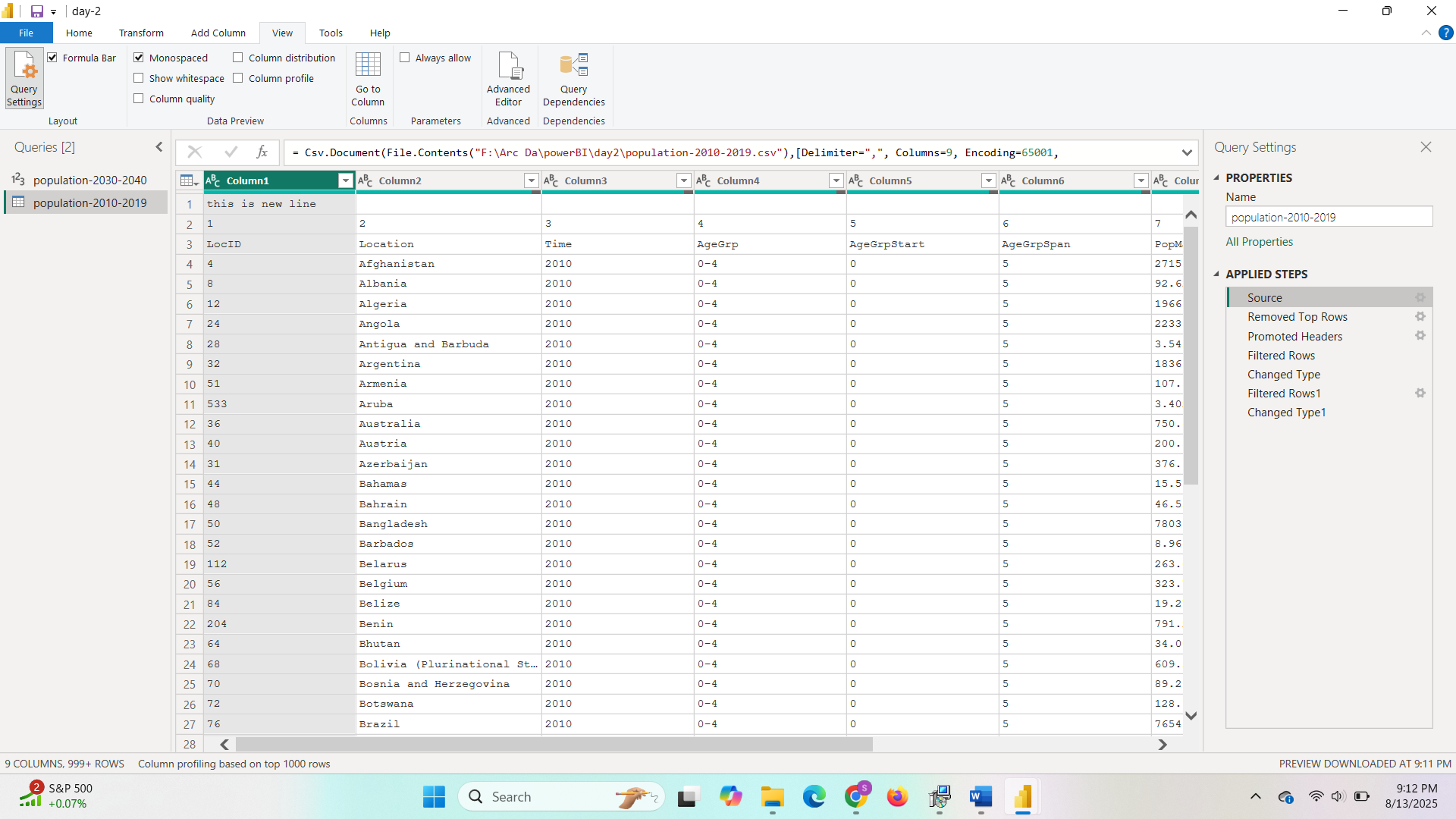
**Data Preparation population 2010-2019**

**Detailed Explanation of Power BI Applied Steps**

The "Applied Steps" pane in Power Query is a chronological record of every cleaning and transformation action you've taken on your data. It's a key feature that makes your work reproducible and easy to audit. Let's break down each step you've highlighted.

**1. Source**

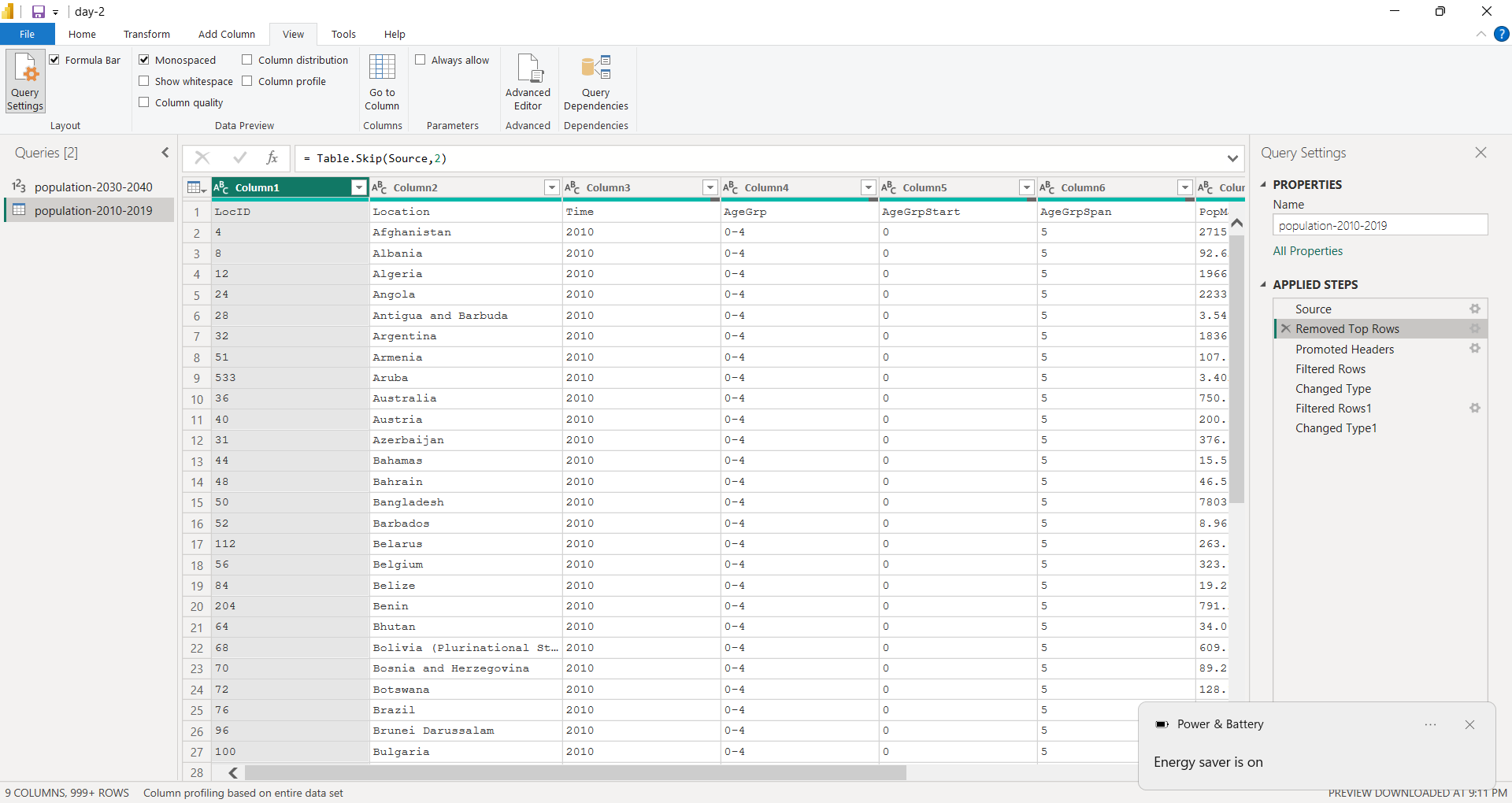
* **What it is:** This is the very first step in your workflow. It represents the connection to your original data source.



* **Detailed Explanation:** In your case, the source is a CSV file. Power Query has loaded the raw data from the file path file:F:\Arc Da\powerBI\day2\population-2010-2019.csv. It also records additional parameters, such as the delimiter (,) and the file encoding (41790), which are essential for correctly parsing the data. This step is a static reference, meaning it points to the original data, so any changes made later in the workflow do not alter the source file. It's the starting point from which all subsequent transformations are applied.

**2. Removed Top Rows**

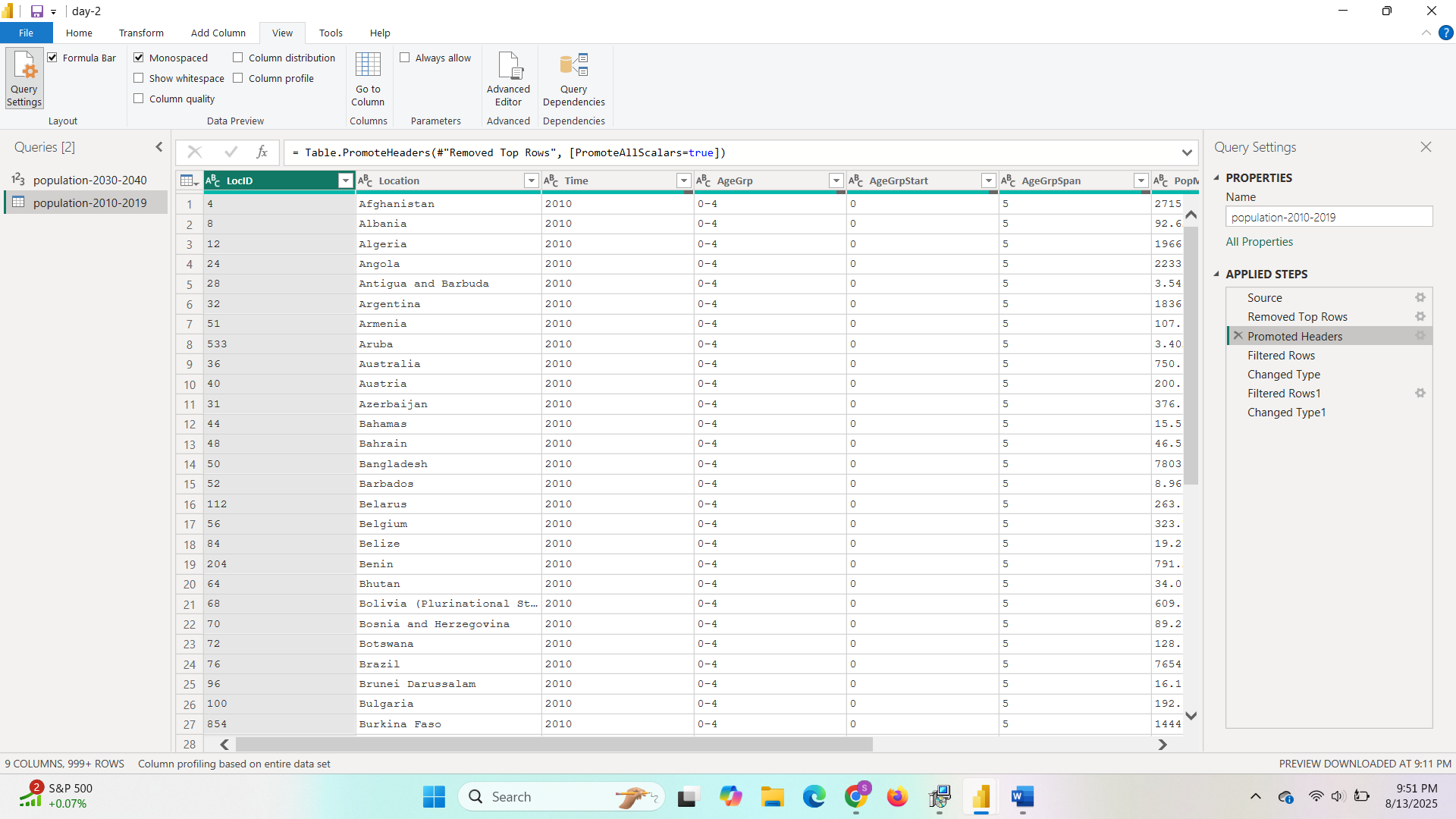
* **What it is:** This step removes a specified number of rows from the top of your dataset.



* **Detailed Explanation:** Looking at your screenshot, the very first row of your raw data contains the text "this is new line"(refer to 1st fig). This is likely a metadata line or an artifact from the source file that is not part of the actual data table. By using the "Remove Top Rows" function, you've deleted this row, so it doesn't interfere with your column headers or data. This is a common practice to remove headers that span multiple rows, descriptive text, or empty space at the beginning of a file.

**3. Promoted Headers**

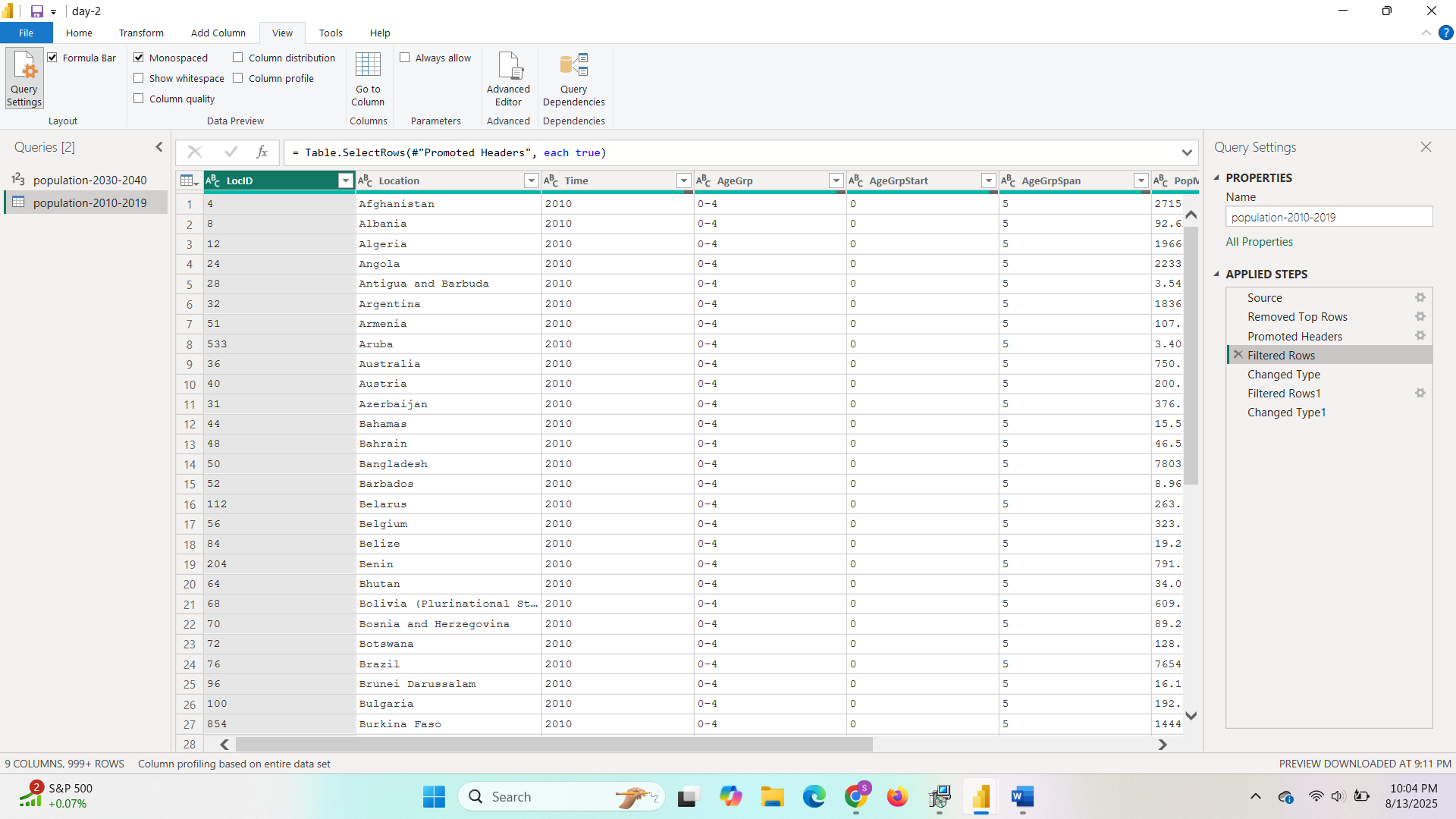
* **What it is:** This step takes the first row of your current table and promotes it to become the column headers.



* **Detailed Explanation:** After removing the top row, the next row (which contains LocID, Location, Time, etc.) is now at the top. This step recognizes that these are your actual column names and sets them as the official headers for your table. This is a fundamental step for making the dataset readable and ready for further transformations. Without this step, your columns would be named Column1, Column2, and so on, which makes data analysis and referencing columns difficult.

**4. Filtered Rows**

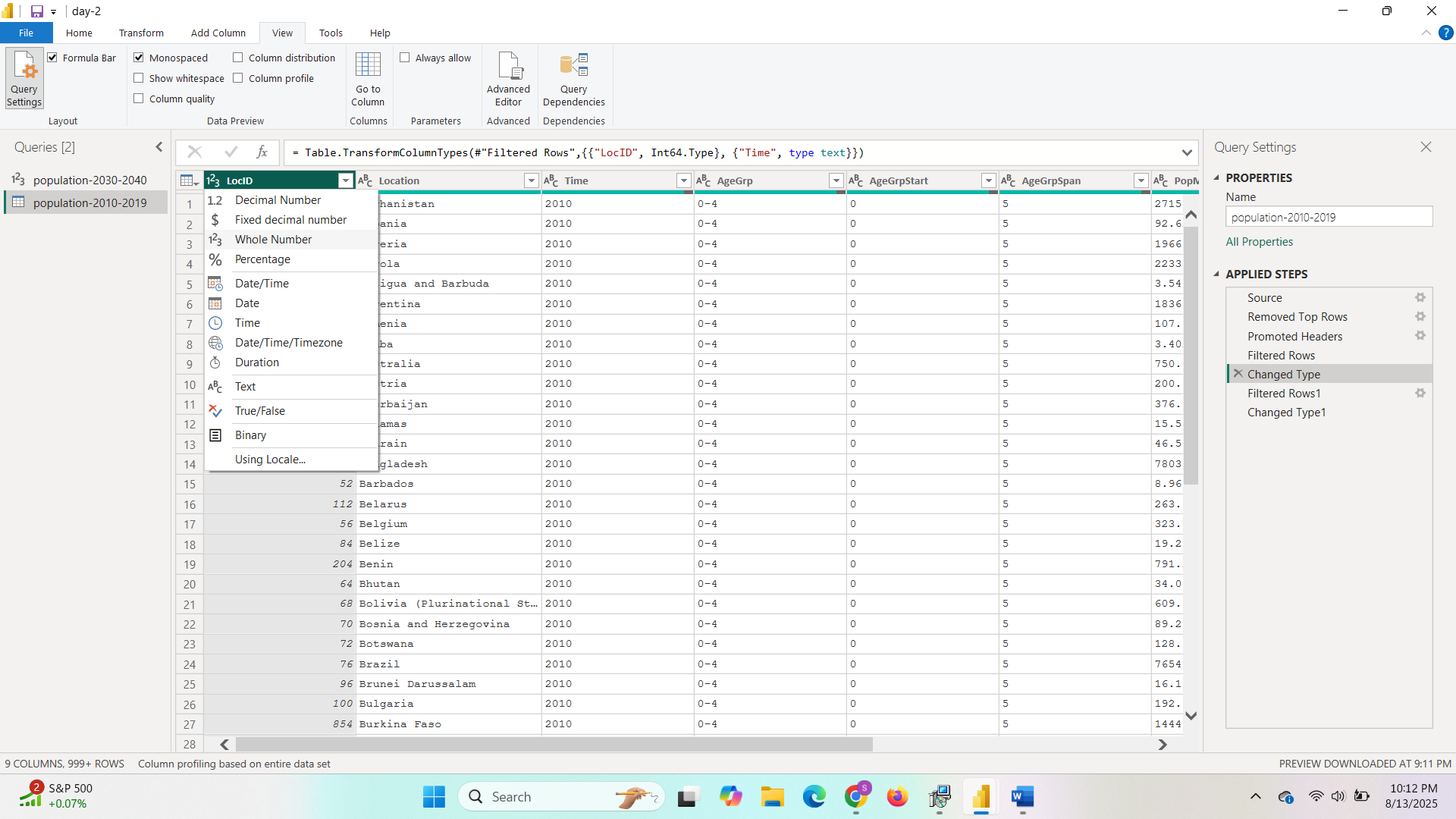
* **What it is:** This step removes rows from the table that do not meet a specific condition.



* **Detailed Explanation:** Based on the common use cases and the order of your steps, this first "Filtered Rows" step was likely used to clean up any remaining rows that might be blank or contain irrelevant information after promoting the headers. For example, if there were empty rows at the beginning or end of your file, this action would remove them to ensure you only have valid data. Power Query allows you to filter based on text values, numerical ranges, or other criteria, giving you precise control over which rows to keep.

**5. Changed Type**

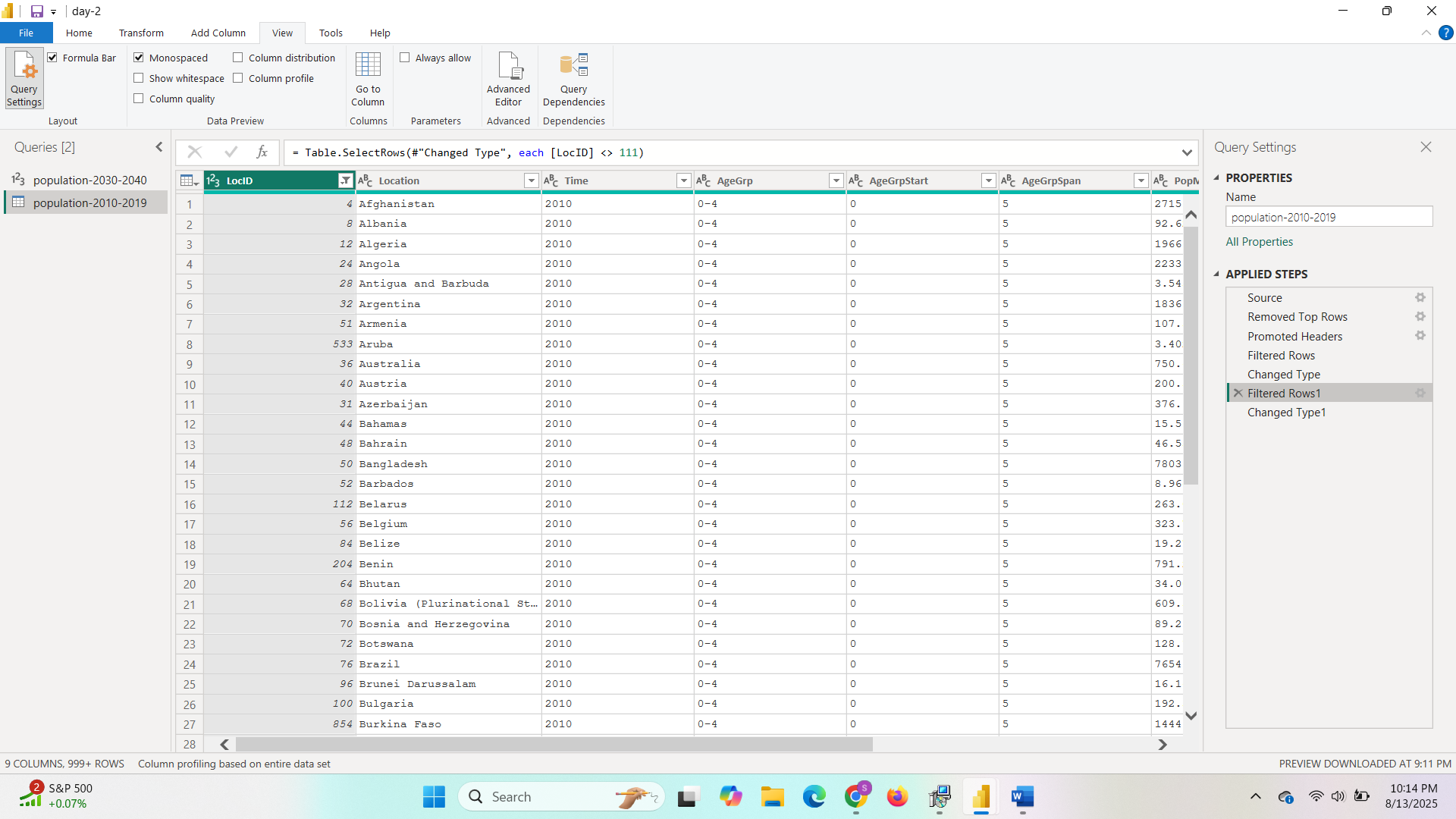
* **What it is:** This step automatically or manually changes the data type of one or more columns.

****

* **Detailed Explanation:** Power Query's "Promoted Headers" action often includes an automatic "Changed Type" step. This is where the editor takes its best guess at the data type for each column (e.g., recognizing 2010 as a whole number, or a country name as text). This is a critical step because the correct data type is essential for performing calculations (e.g., you can't sum a text column) and for creating relationships between tables. It's a good practice to always review this step and manually adjust any types that were not correctly identified.
  + Time (Year): Whole Number
  + AgeGrpStart : Whole Number
  + AgeGrpSpan : Whole Number
  + PopMale : Decimal

**6. Filtered Rows (Second Instance)**

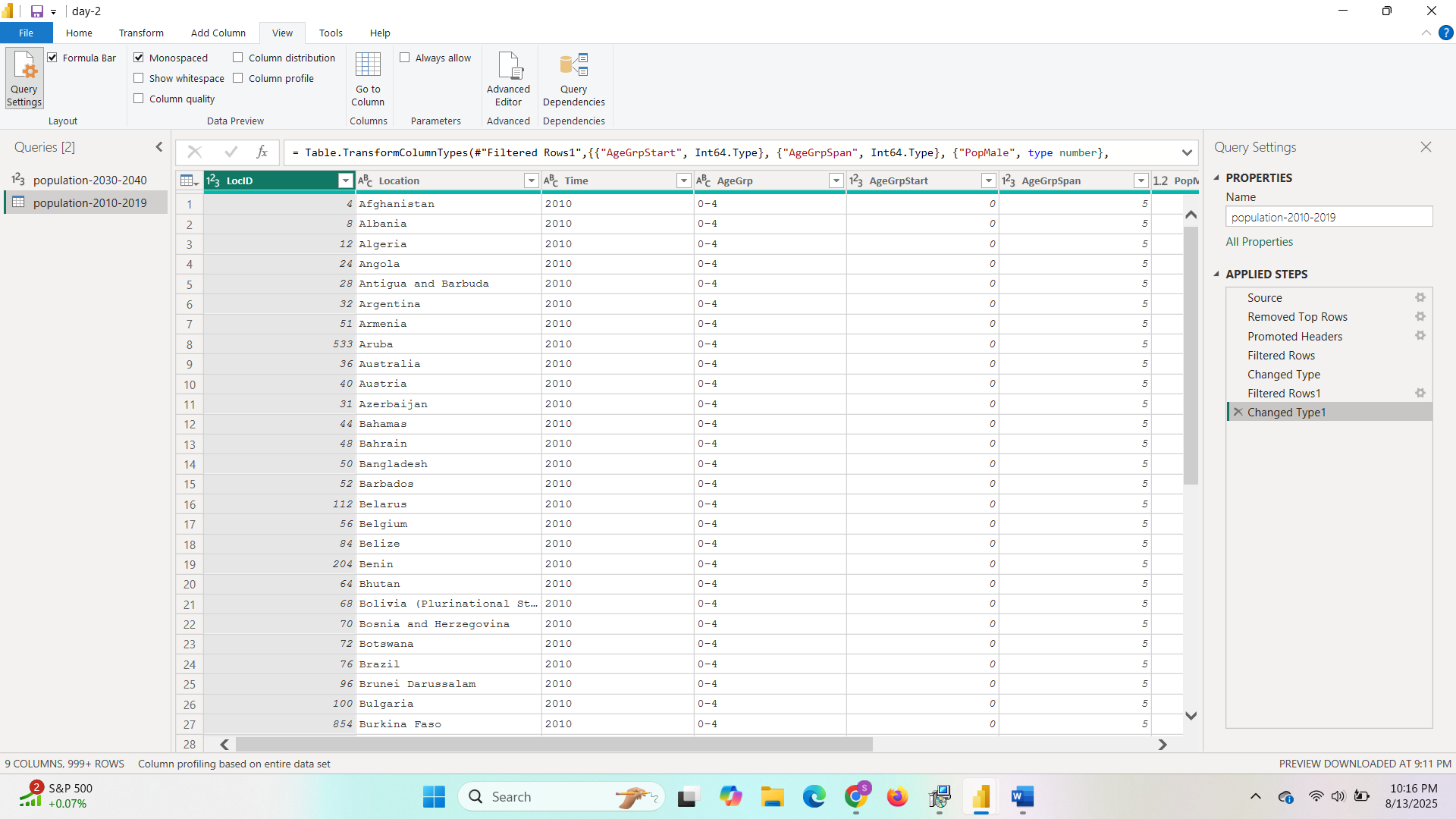
* **What it is:** This is another instance of filtering, which you have performed again later in the process.

****

* **Detailed Explanation:** After the initial cleaning and data type conversion, you likely noticed some additional rows that you wanted to exclude from your analysis. The first filter might have been a general cleanup, while this second filter is more likely a specific analytical choice. For example, you may have decided to only include data for specific countries, a particular time range, or a certain age group. This highlights the iterative nature of the data cleaning process. Targeted dataset enables accurate analysis, reduces noise, and boosts performance in Power BI visuals.

**7. Changed Type1**

* **What it is:** This is a second data type change step.



* **Detailed Explanation:** Power Query automatically appends a number to a step name when you perform the same action more than once. In this case, "Changed Type1" indicates that you manually went back and changed the data type of one or more columns again, perhaps to correct an earlier mistake or to reformat a column (e.g., converting a text field to a date after the initial type change). This can also happen after a transformation that changes the output data type. This step reinforces the fact that you can always go back and make changes to your workflow, and Power Query will re-apply the steps in the correct order.

# The Final View in Power Query

The view you're looking at is the final output of your data preparation workflow in the Power Query Editor. This is what your dataset will look like when it's loaded into the main Power BI desktop for building reports and visualizations. This view also provides key insights into the quality and structure of your prepared data.

**Column Profiling**

At the top of the data preview, you'll see a series of visual bars and metrics. This is Power Query's **Column Profiling** feature, which is an invaluable tool for understanding your data.

* **Valid / Error / Empty**: This section gives you a quick visual summary of the quality of your data for each column.
  + **Valid**: The percentage of cells in the column that contain valid, correctly formatted data.
  + **Error**: The percentage of cells that contain errors (e.g., text in a column that was designated as a number).
  + **Empty**: The percentage of cells that contain null or empty values.
  + In your screenshot, the LocID, Location, and Time columns show **100% Valid** and **0% Empty**, which tells you that your previous cleaning steps were successful for these columns.
* **Distinct / Unique**: These metrics give you a sense of the cardinality of your data.
  + **Distinct**: The number of different, non-repeating values in a column.
  + **Unique**: The number of values that appear only once in the column.
  + For the Location column, you can see 199 distinct values, which likely corresponds to the number of countries or locations in your dataset.

