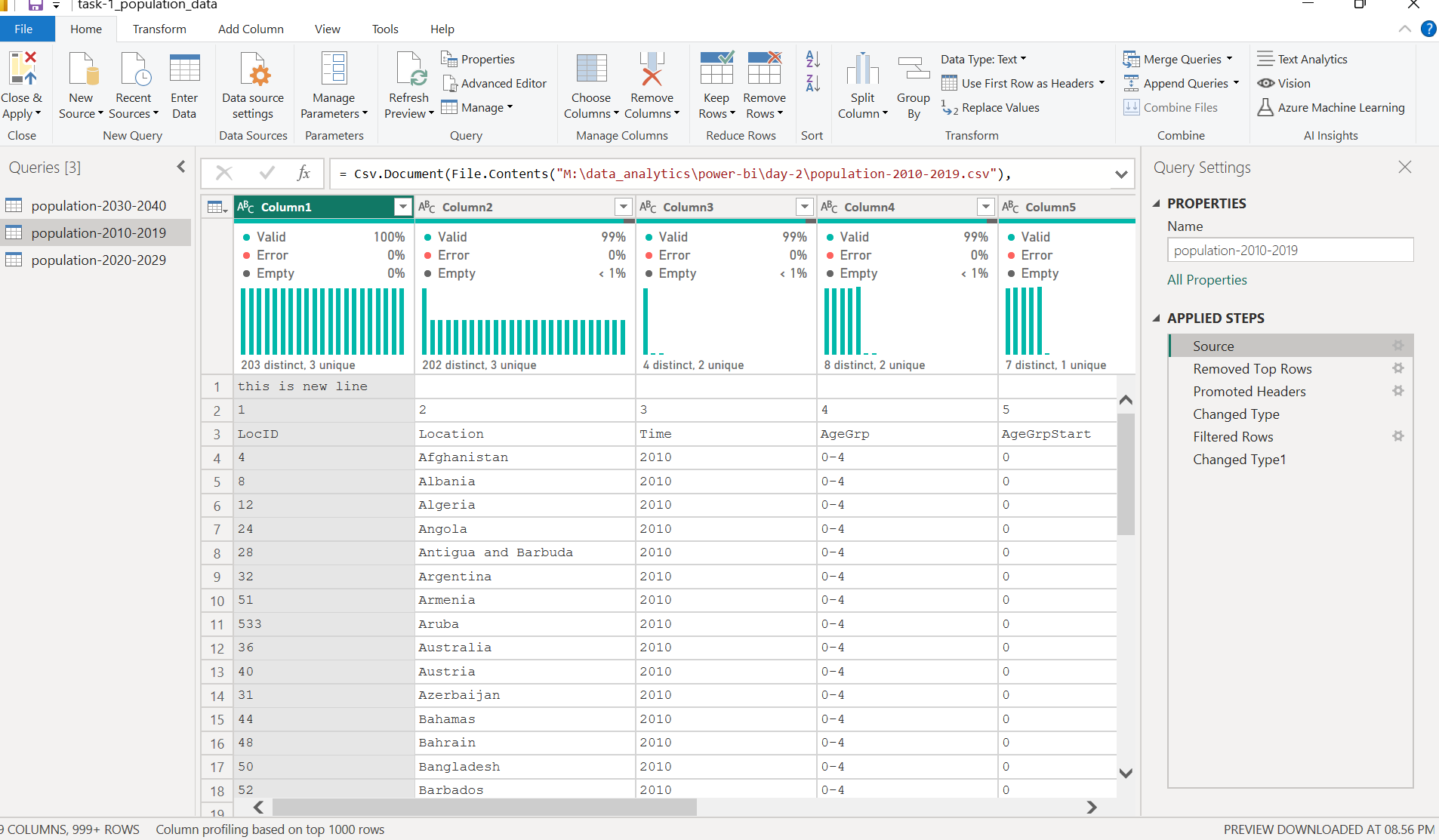
# Data Cleaning, Transformation, and Documentation Workflow

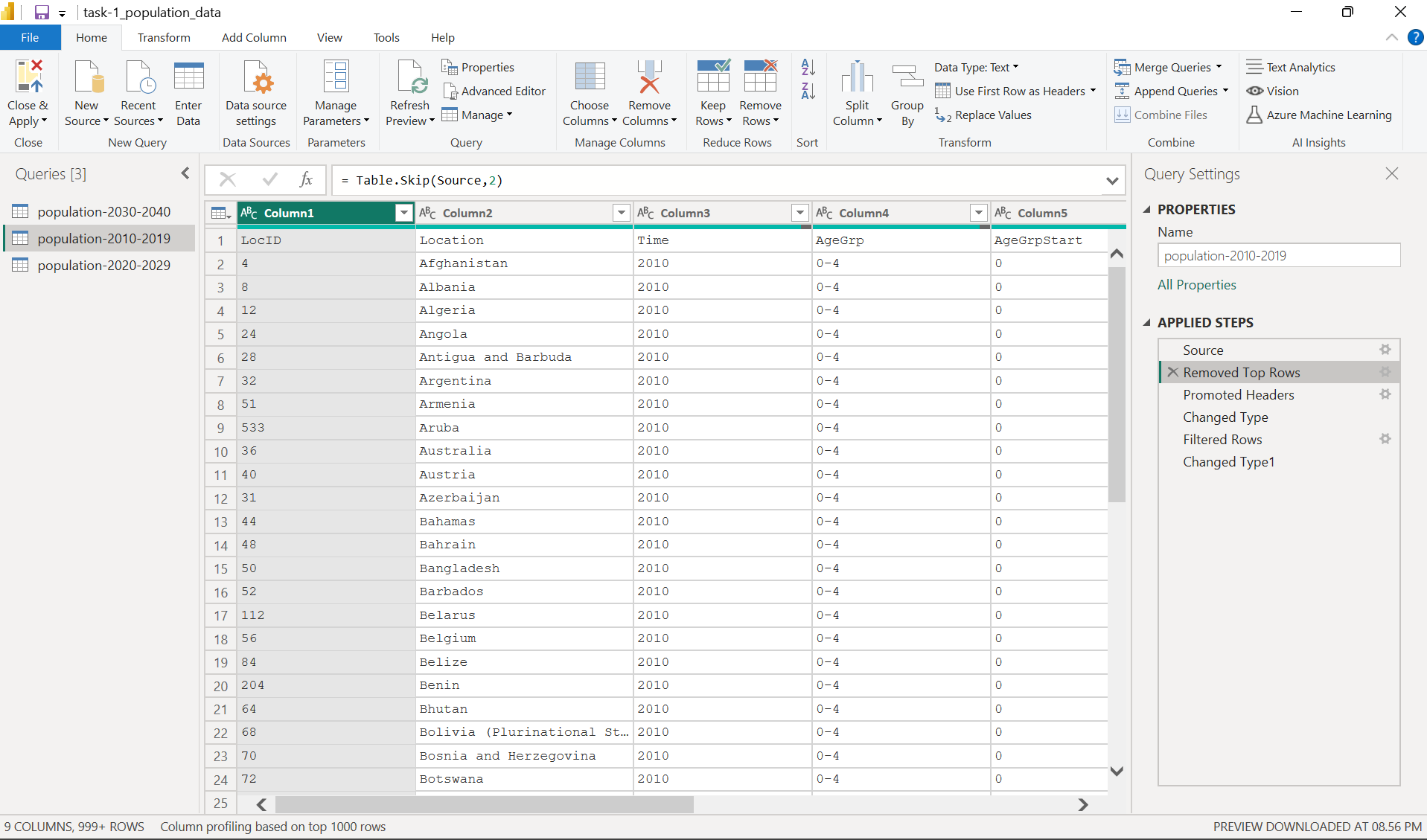
**Data Cleaning Workflow**

This workflow focuses on preparing the data for analysis by addressing common issues.

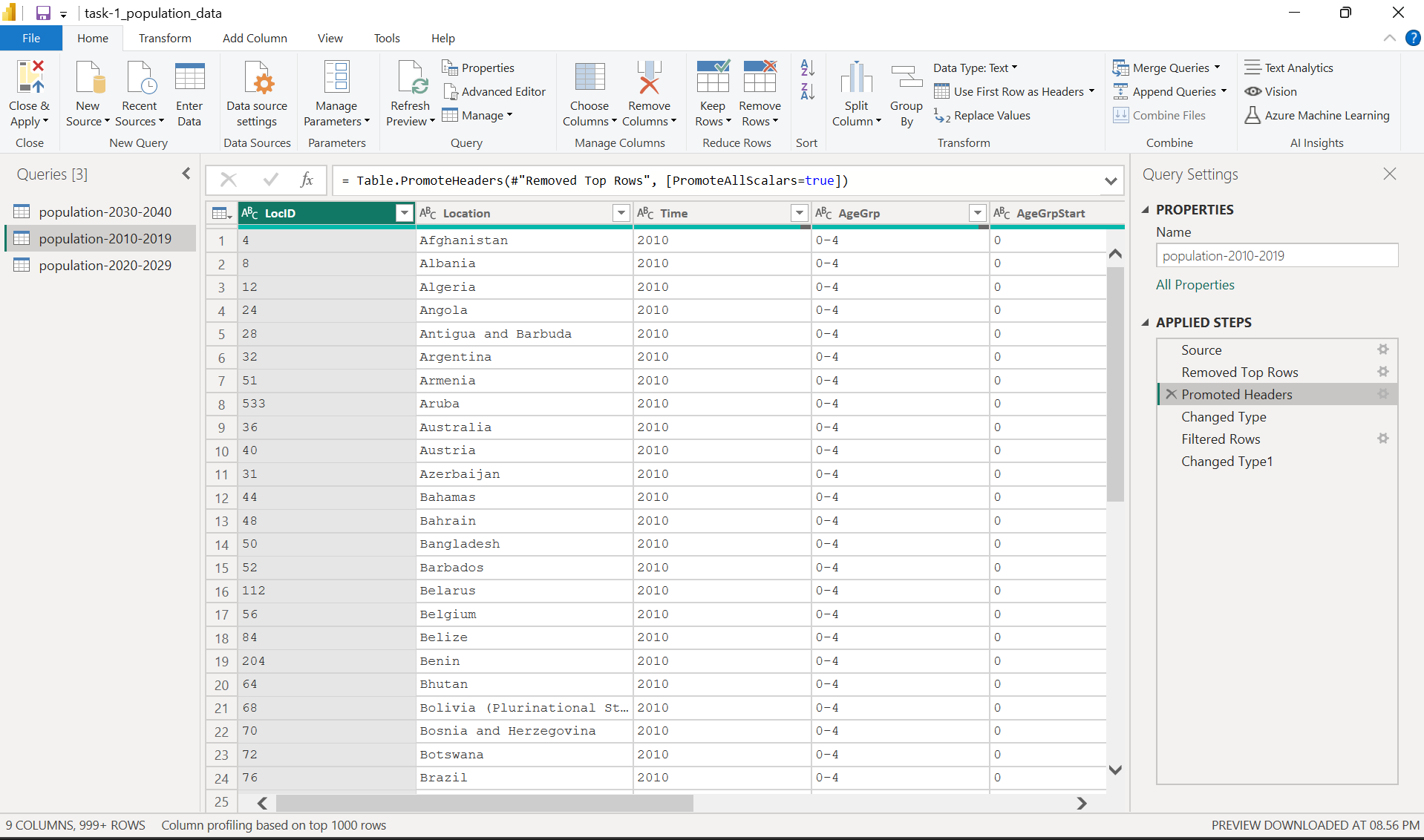
1. **Source:** This first step represents the connection to your raw data, such as a spreadsheet or database. It loads the data in its original, untouched state into the Power Query Editor.



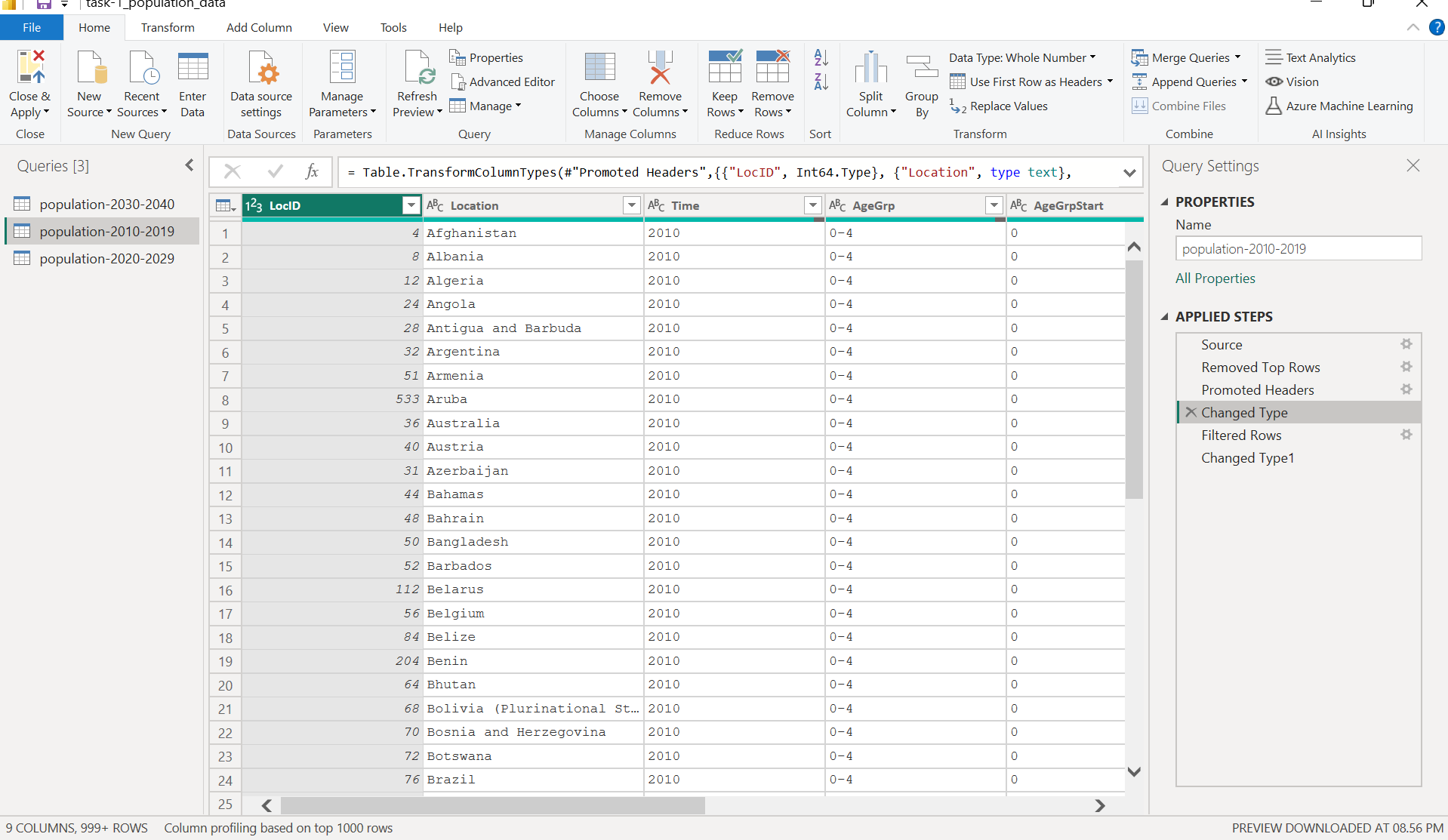
1. **Removed Top Rows:** This is a crucial cleaning step. The image shows that some rows were removed from the top of the dataset. This is typically done to eliminate non-data information, like report titles or blank lines, that are often present in raw files.



1. **Promoted Headers:** This step takes the first meaningful row of data and sets it as the column headers. This is why the columns are correctly labeled as **Location**, **Time**, and **AgeGrp** instead of generic names.



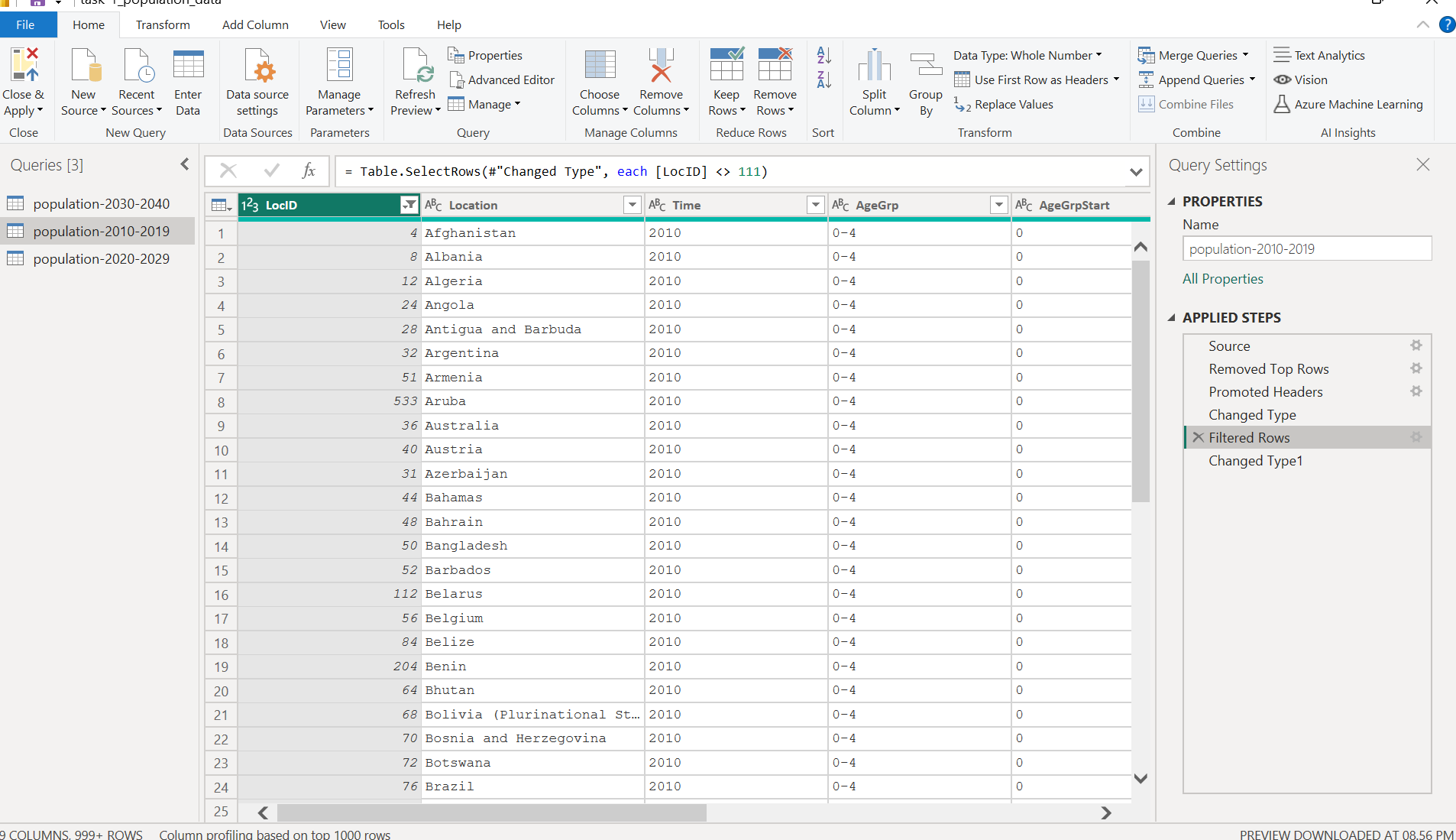
1. **Changed Type:** This step is about setting the correct data type for each column. For example, a column containing years would be set to a numerical type, while a column with country names would be set to text. This is fundamental for accurate sorting and calculations.



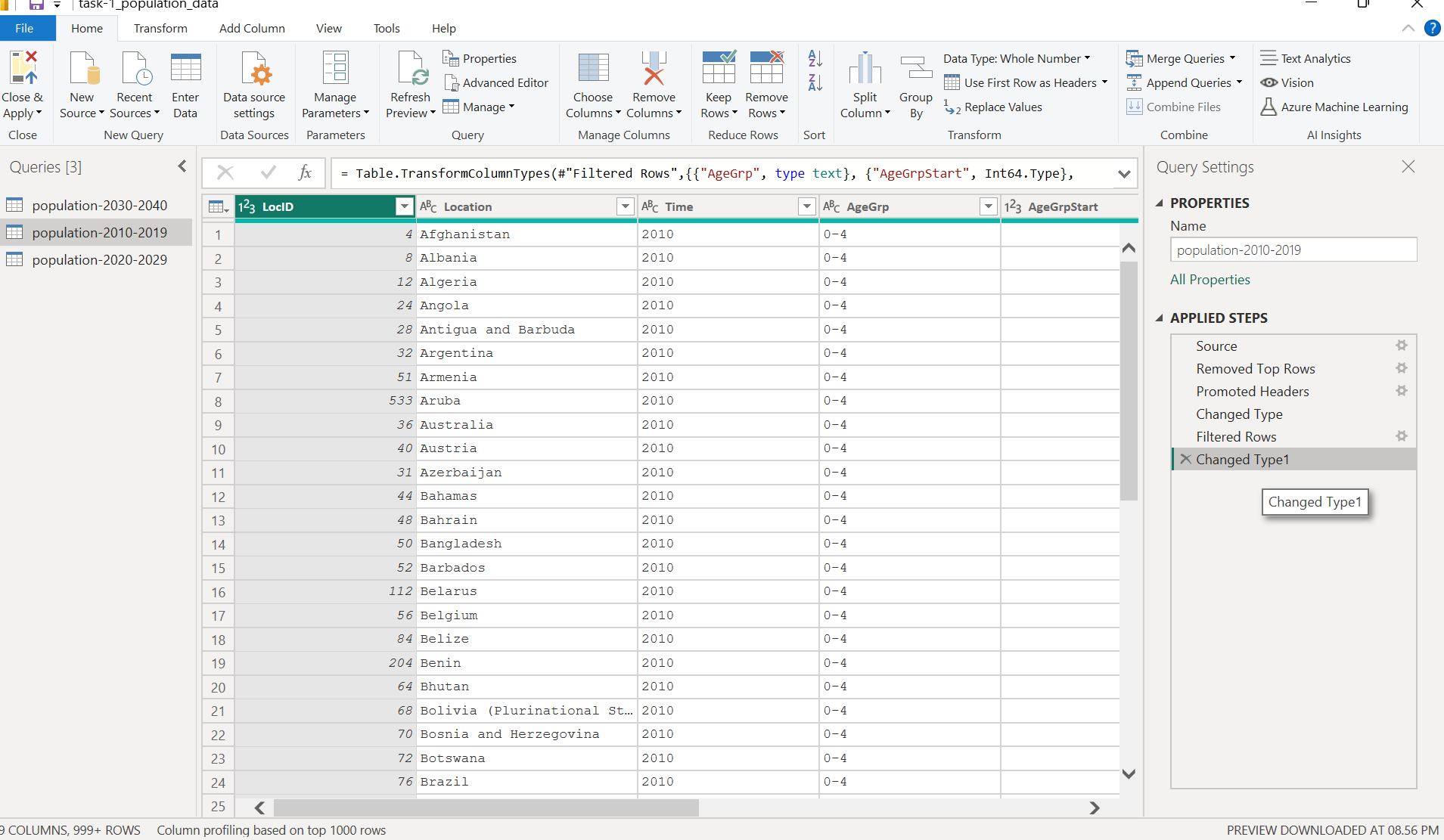
**Data Transformation Workflow**

This workflow involves modifying the data's structure and content to make it suitable for a specific analysis.

1. **Filtered Rows:** The image shows a Filtered Rows step. The query name, population-2010-2019, strongly suggests that a filter was applied to the **Time** column to include only data from a specific time range. This reduces the dataset to a manageable size and focuses the analysis on the relevant data.



1. **Changed Type1:** This step represents a second data type adjustment. It's common to make these adjustments after other transformations, like filtering, to ensure the data types are still correct. The image shows that the **AgeGrp** column was set to text and **AgeGrpStart** was set to an integer (Int64.Type).



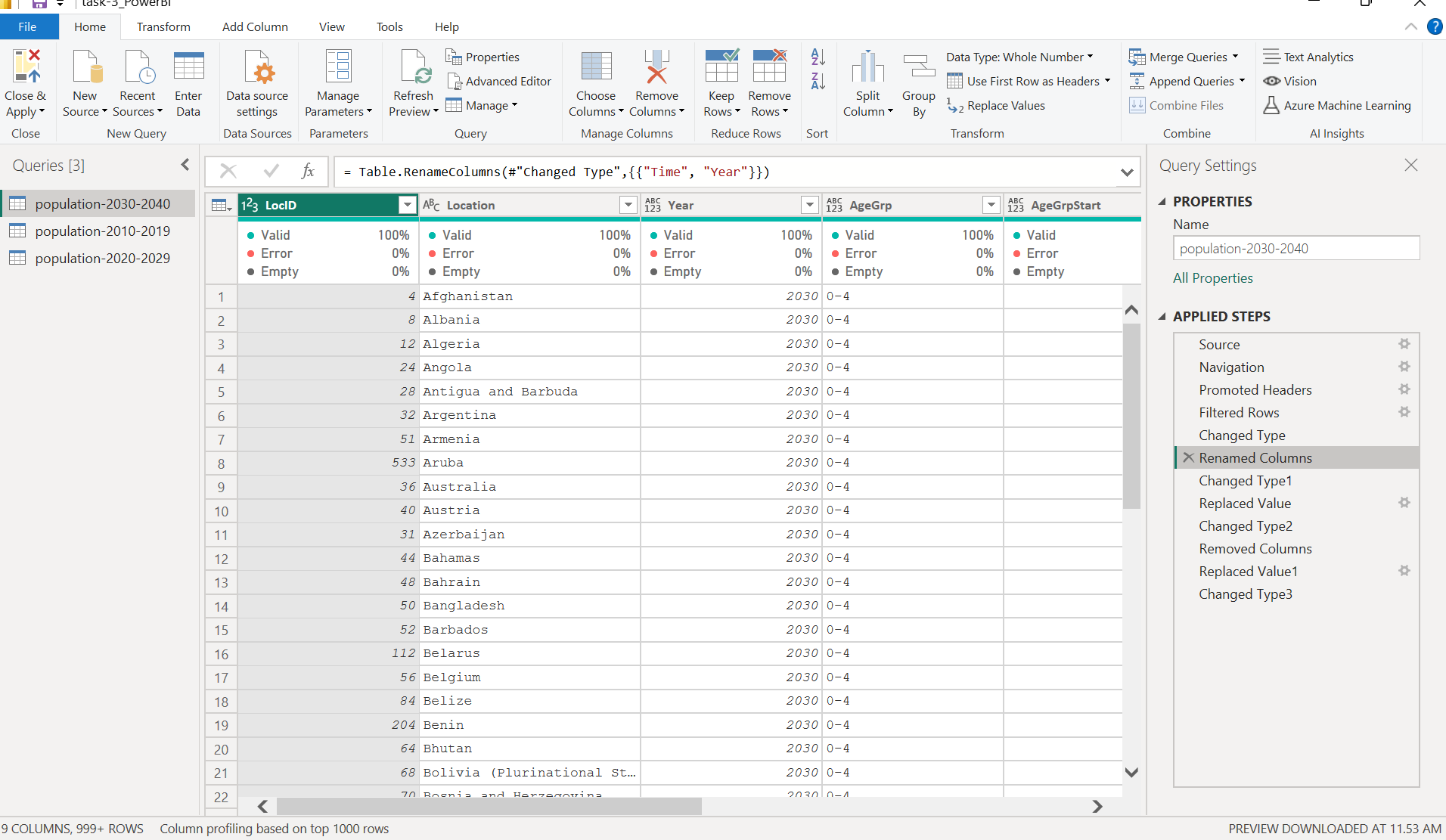
**Documentation Workflow**

The Power Query Editor has built-in features that automatically document your entire workflow.

1. **Applied Steps:** This panel, located on the right side of the screen, automatically records every cleaning and transformation action you perform. It creates a complete history of your work, making the process transparent and reproducible. You can click on any step to see what the data looked like at that point.
2. **Query Name:** A descriptive name, like population-2010-2019, is a simple but effective form of documentation. It immediately tells you and others what the query contains, making it easy to identify its purpose without having to examine the data.

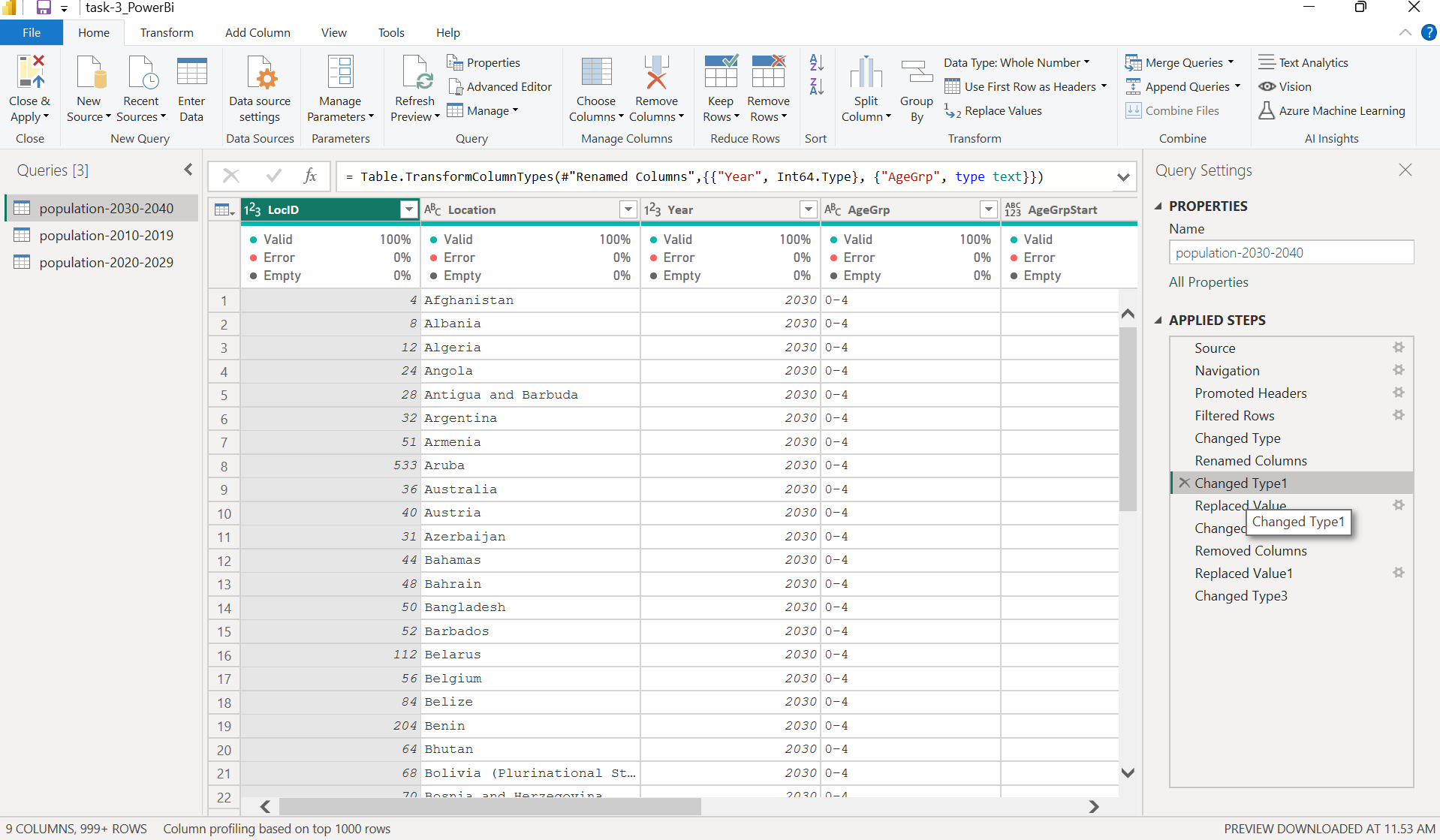
**5. Renamed Columns:**

* **Explanation:** This is a manual step where one or more column headers have been renamed. The image shows that a column, likely containing year information, has been explicitly renamed to "Year" to make the dataset more understandable.



**6. Changed Type1:**

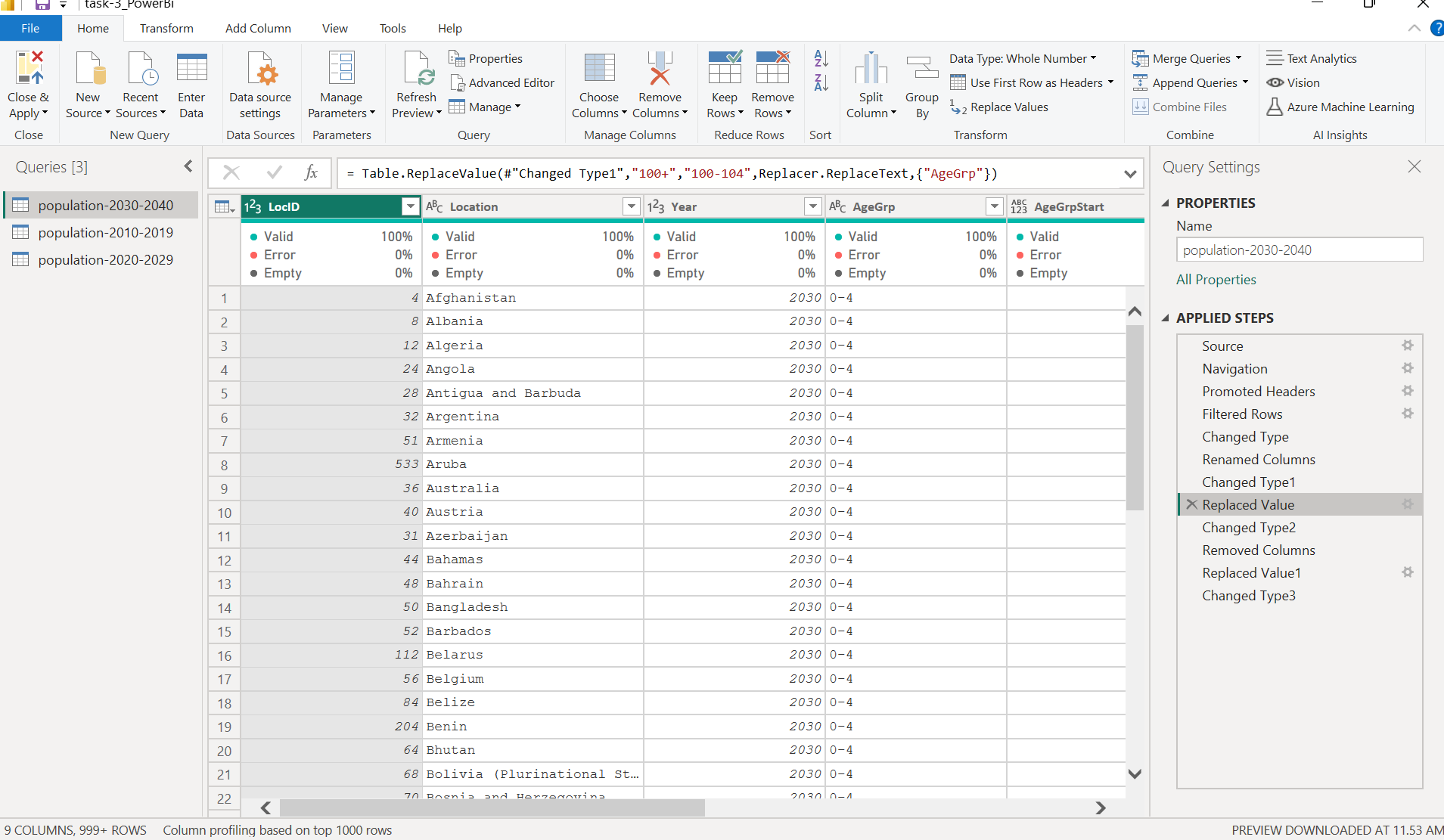
* **Explanation:** This is a subsequent data type change. After a column was renamed (Step 5), the data type was likely adjusted again to ensure it is correct for analysis. For instance, the "Year" column might have been changed to a "Whole Number" type.



**7.Replaced Value**

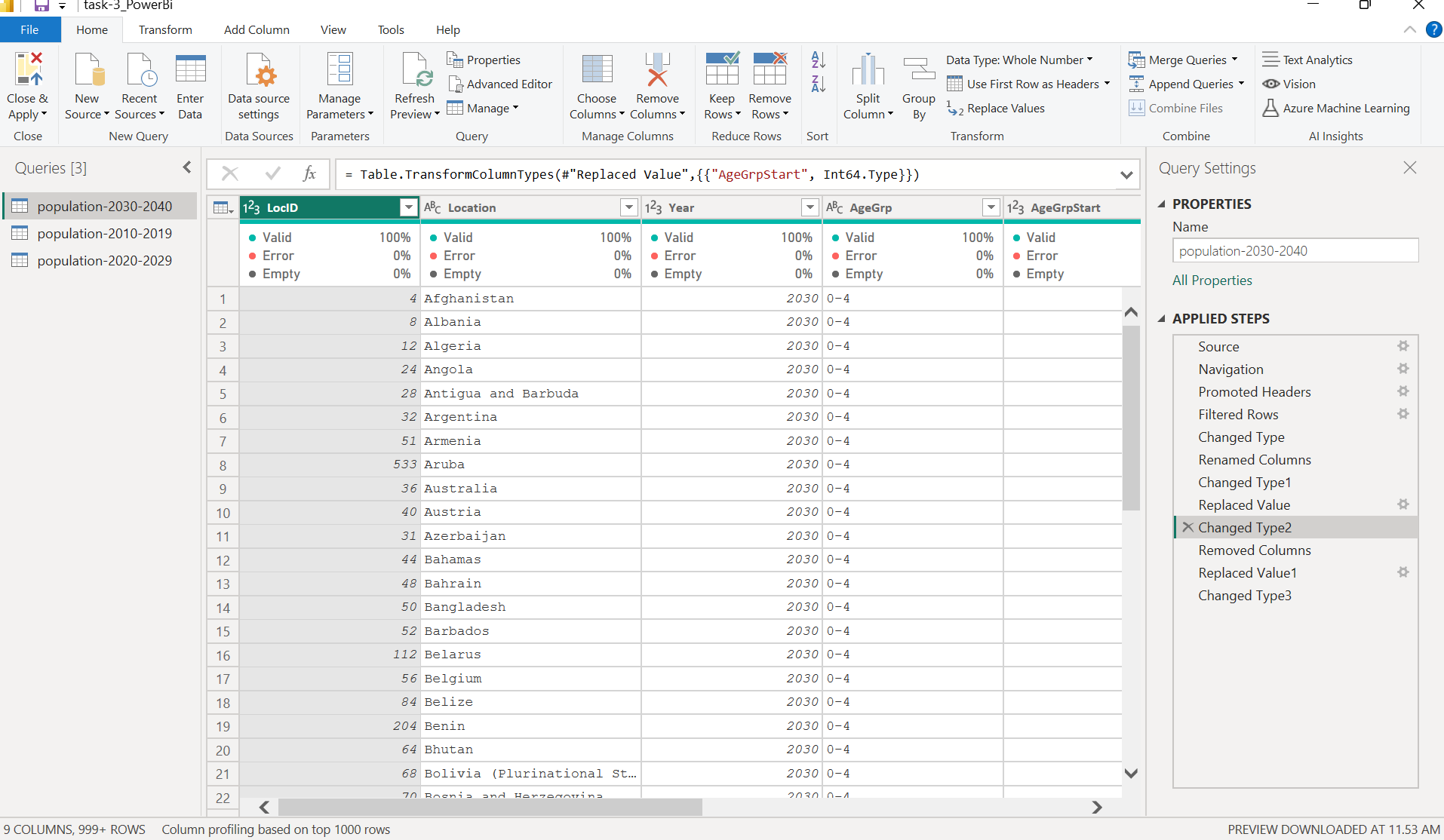
**This step is the first transformation applied after the columns were renamed.**

* Explanation: This indicates that a specific value or set of values within a column was found and replaced with a new value. This is a common data-cleaning task, often used to correct typos, standardize entries, or change a specific code to a more readable name.



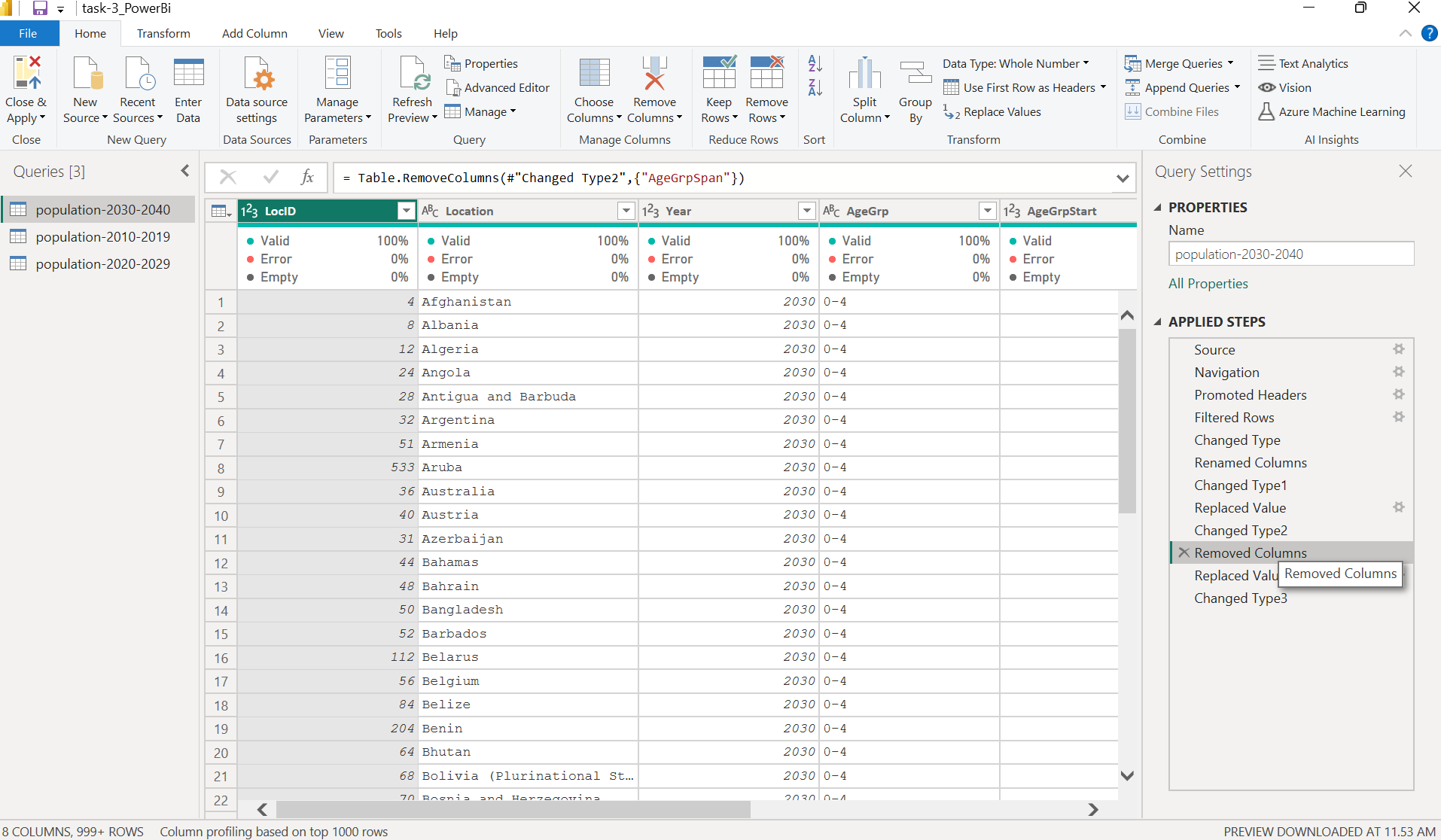
**8. Changed Type2:**

* **Explanation:** This indicates a second, distinct set of data type changes applied to the dataset. It could be that additional columns were identified and their data types were corrected after the previous steps.



**8. Removed Columns:**

* **Explanation:** This step shows that one or more columns from the dataset have been removed because they were not needed for the analysis. This helps to reduce the size and complexity of the data.



**9. Replaced Value1:**

* **Explanation:** In this step, a specific value or set of values within a column has been replaced with a new value. This is often done to standardize data entries or correct errors, such as replacing a misspelled location name or a specific code with a more descriptive name.

