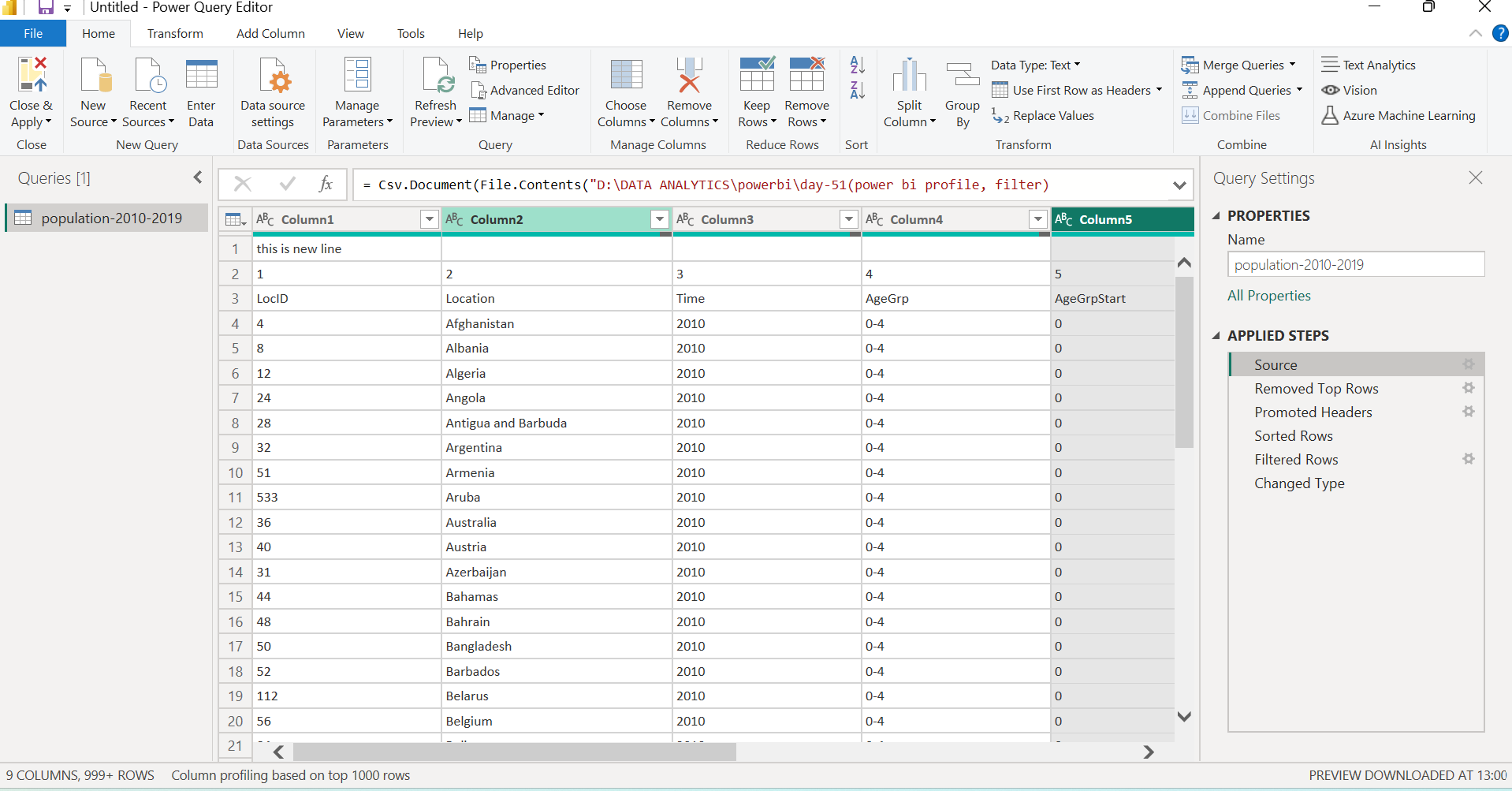
**Data Preparation Explained: Power BI (Population\_2010-2019)**

Data import in Power BI is the essential first step in the data analysis workflow. It's the process of connecting to a data source and loading the data into your Power BI Desktop file. This action creates the foundation for all subsequent data modeling, analysis, and report creation.

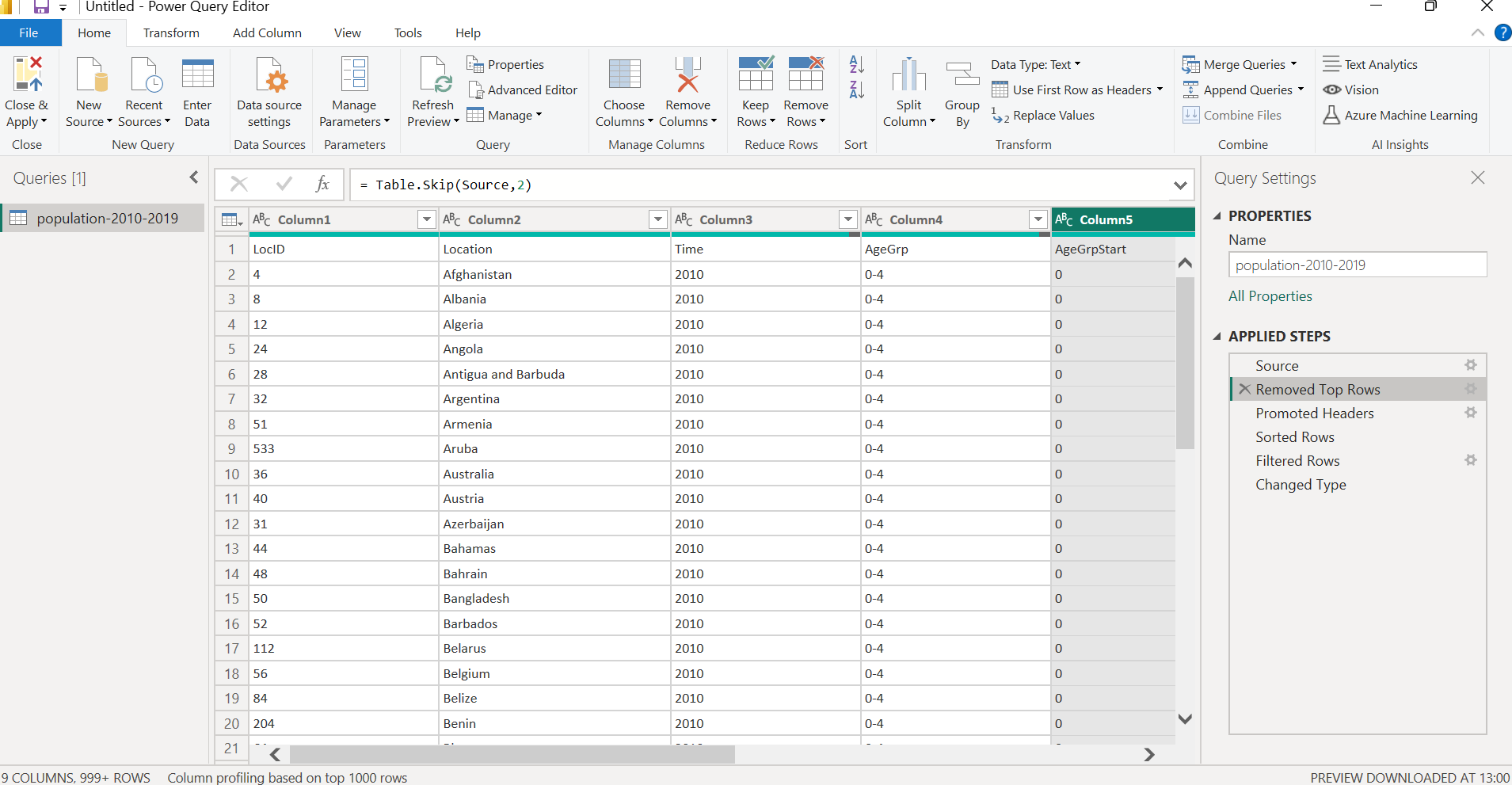
**The Data Import Process:**

The process of importing data into Power BI Desktop is straightforward:

1. **Select "Get Data":** On the **Home** tab of Power BI Desktop, click on the **Get Data** button. This opens a window showing a list of common data sources. You can also select the "More..." option to see the full list of over 100 supported data sources. 2. **Choose Your Data Source:** Power BI categorizes data sources into several groups:
   * **File:** This includes common file types like Excel workbooks, CSV files, and PDFs.
   * **Database:** This category contains connectors for popular databases such as SQL Server, Oracle, and MySQL.
   * **Power Platform:** This allows you to connect to data from other Power Platform services like Power BI dataflows and Dataverse.
   * **Azure:** This group contains connectors for various Azure services, including Azure SQL Database and Azure Synapse Analytics.
   * **Online Services:** This includes connections to services like SharePoint Online, Google Analytics, and Salesforce.
   * **Other:** This is a catch-all for web pages, ODBC, and other sources.
2. **Connect and Configure:** After selecting a data source, you'll be prompted to provide the necessary connection details, such as a file path, server name, or URL, along with any required credentials.
3. **Load or Transform:** Once the connection is established, the **Navigator** window appears. This window allows you to preview the data and select the tables you want to import. At this point, you have two options:
   * **Load:** This option directly imports the selected data into your Power BI data model.
   * **Transform Data:** This opens the **Power Query Editor**, where you can clean and shape the data before loading it. This is a best practice for most data import scenarios.

**Removed Top Rows:**

To remove top rows in Power BI, you must use the **Power Query Editor**. This is a common step when your data source, such as an Excel file or CSV, contains introductory information or empty rows before the actual data table begins.

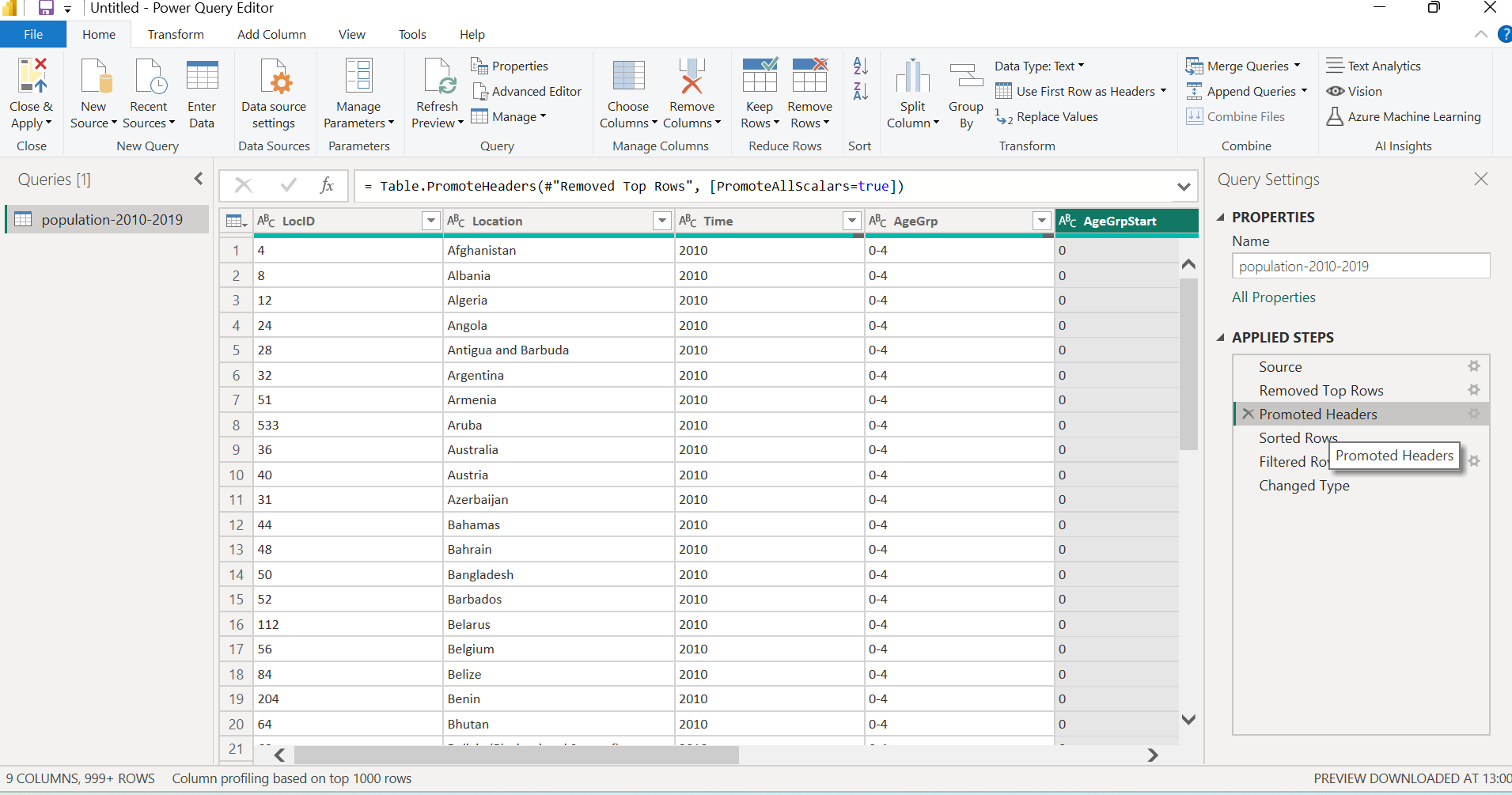


**Step-by-Step Guide**

1. Open the **Power Query Editor**. You can do this by clicking **Transform data** on the **Home** tab of Power BI Desktop.
2. Navigate to the query (table) you want to modify in the **Queries** pane on the left.
3. On the **Home** tab of the ribbon, find the **Reduce Rows** group.
4. Click on **Remove Rows**.
5. From the drop-down menu, select **Remove Top Rows**.
6. A dialog box will appear. Enter the **Number of rows** you wish to remove from the top of the table. For example, if your table has two header rows before the data begins, you would enter "2".
7. Click **OK**. The rows will be removed, and a new step named "**Removed Top Rows**" will appear in the **Applied Steps** pane on the right.

**Promoted Headers:**

To "promote headers" in Power BI is to transform the **first row of data into the column names** for your table. This is a crucial step in data cleaning, as many raw data sources (like CSV or Excel files) don't have properly formatted headers.



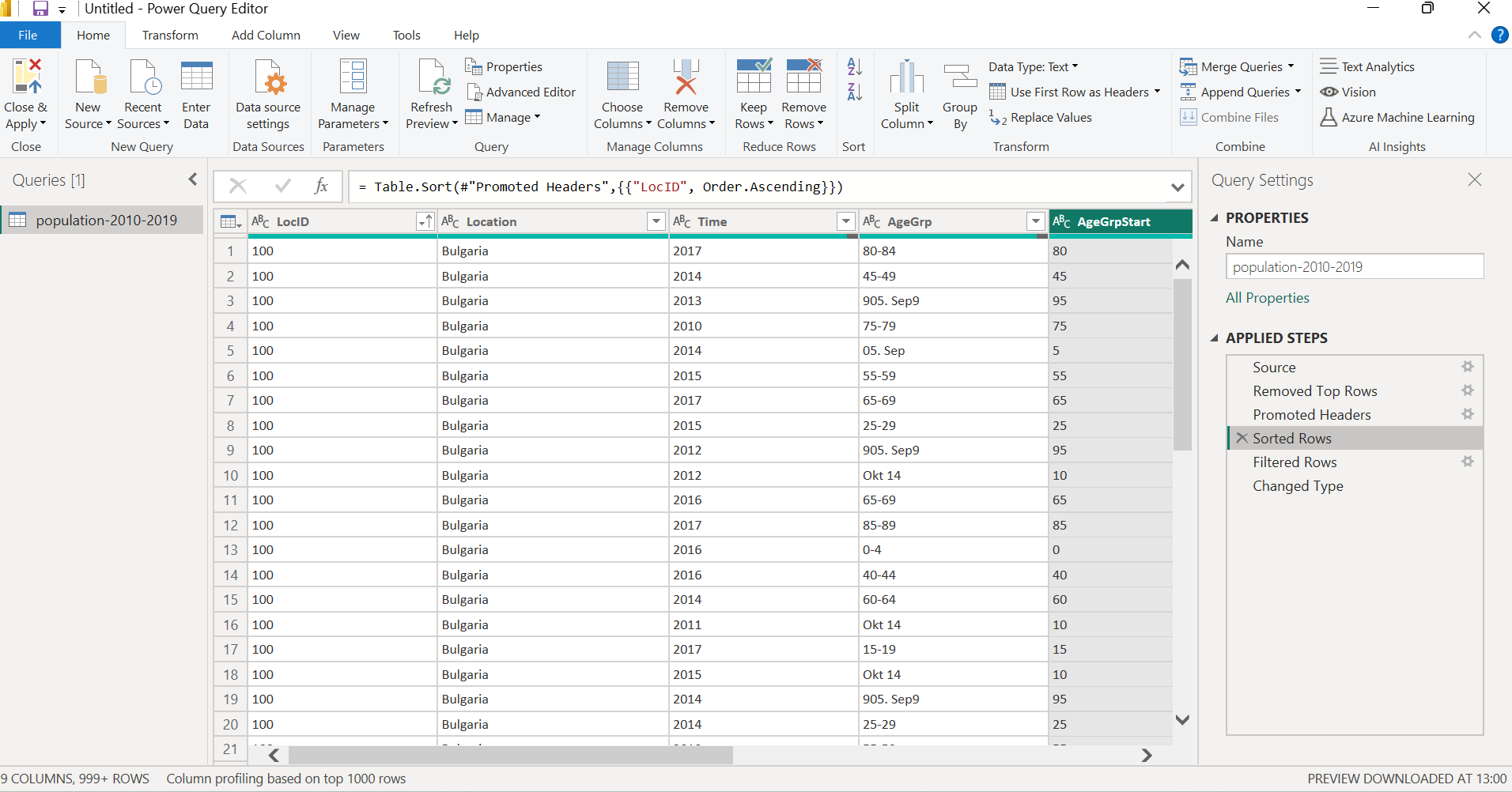
**How to Promote Headers**

You perform this action in the **Power Query Editor**. It's a simple, one-click process.

1. Open the **Power Query Editor** by clicking **Transform data** in Power BI Desktop.
2. Select the table you want to clean.
3. On the **Home** tab, in the **Transform** group, click on the **Use First Row as Headers** button.

Power Query will automatically move the values from the first row to the column header row. It will then also add a new **Applied Step** called **Promoted Headers** to the list on the right side of the screen.

**Sorted Rows:**

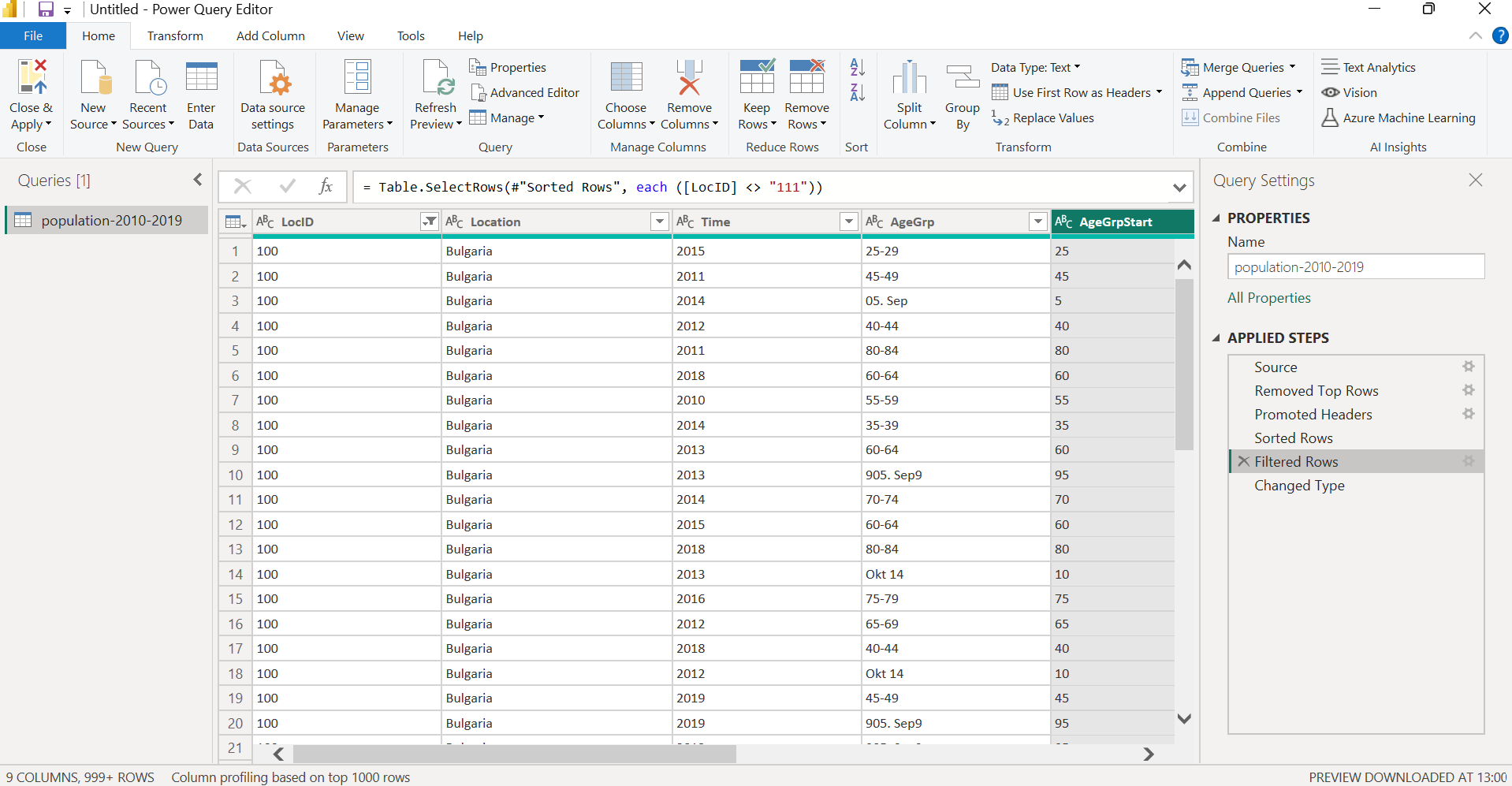


The quickest way to sort data is directly in the **Power Query Editor**. This permanently changes the order of the rows in your table.

1. Open the Power Query Editor by clicking **Transform data** on the Home tab.
2. Find the column you want to sort.
3. Click the drop-down arrow in the column's header.
4. Select **Sort Ascending** or **Sort Descending**.

Power Query adds a new **Applied Step** called "**Sorted Rows**," making the sort a repeatable part of your data preparation.

**Filtered Rows:**

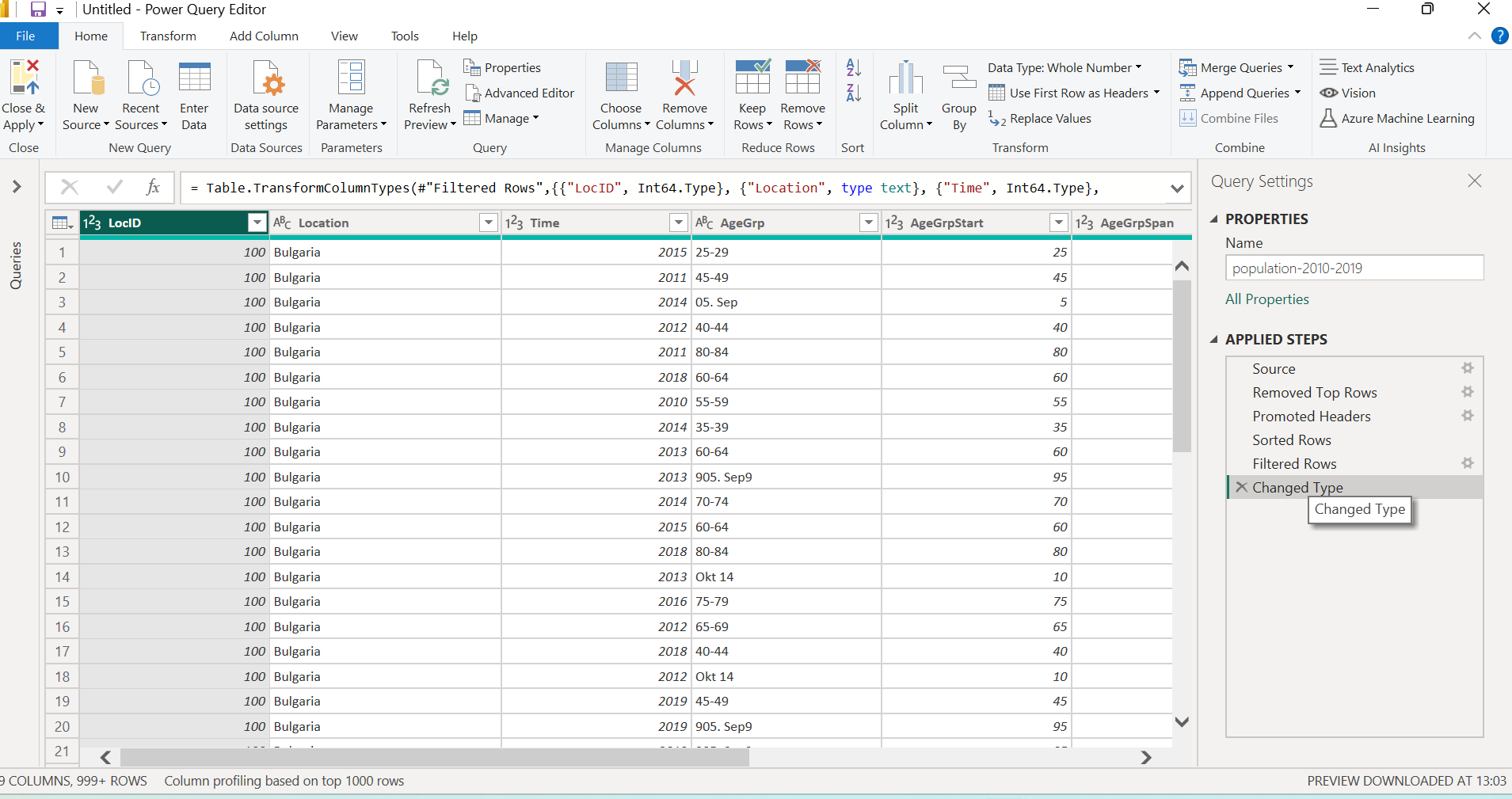
****

You can apply filters in two main ways:

* **Filter by Value**: You click the dropdown arrow on a column header and select or deselect specific values to keep or remove. For example, you could filter a "City" column to only show "New York" and "London."
* **Advanced Filtering**: This allows you to set more complex conditions, such as filtering a "Sales" column to keep only values "greater than 1000" or a "Date" column to show only sales "after a specific date."

Each time you apply a filter, a new step called "**Filtered Rows**" is added to the **Applied Steps** pane in the Power Query Editor. This makes the filtering a permanent and repeatable part of your data cleaning process. When you refresh your data, the same filtering rules are automatically applied again.

**Change Data Type:**



Changing a column's data type in Power BI means specifying the kind of data it contains, such as text, numbers, or dates. This is a crucial data preparation step because it ensures that Power BI interprets your data correctly, allowing for accurate calculations and visualizations.

**Why is it Important?**

* **Correct Calculations**: If a column of numbers is incorrectly identified as text, you can't perform mathematical operations like summing or averaging.
* **Accurate Visualizations**: Data types dictate how visuals are rendered. For example, a date column allows you to create a time-based chart, while a text column would not.
* **Improved Performance**: Power BI can work more efficiently with properly defined data types, especially for numerical and date data, which take up less space.