**Steps to Access Power Query in Excel**

1. **Open Excel**:
   * Open Microsoft Excel.
2. **Load Your Dataset (87508 rows including the header row and 8 columns)**:
   * Go to the File tab.
   * Select Open and choose your audible\_uncleaned.csv file.
3. **Access Power Query**:
   * Go to the Data tab on the Excel ribbon.
   * Look for the Get & Transform Data group.
   * Click on Get Data > From File > From Text/CSV.
4. **Load the CSV into Power Query**:
   * Select your audible\_uncleaned.csv file and click Import.
   * In the Navigator window, select the sheet or table you want to import.
   * Click Transform Data to clean the data in Power Query Editor.

**Using Power Query to Clean the Data**

1. **Remove Prefixes from author and narrator Columns**:
   * Select the author column.
   * Go to the Transform tab.
   * Click on Replace Values.
   * In the Value to Find box, type Writtenby:.
   * Leave the Replace with box empty and click OK.
   * Repeat the same steps for the narrator column, replacing Narratedby: with an empty string.
2. **Add Spaces between Author and Narrator Names**:
   * **Split Columns**:
     + Select the author column.
     + Go to the Transform tab.
     + Click on Split Column > By Lowercase to Uppercase.
     + This will split the column at the transition from lowercase to uppercase letters, effectively separating the first, middle, and last names.
   * **Merge Columns with Space**:
     + **For author columns**:
       - Select all the split author columns (e.g., author.1, author.2, etc.).
       - Go to the Transform tab.
       - Click on Merge Columns.
       - In the Merge Columns dialog, select Space as the separator.
       - Name the new column (e.g., Cleaned Author) and click OK.
     + **For narrator columns**:
       - Repeat steps above for the narrator columns to merge with a space separating each name.
3. **Remove Duplicates**:
   * Select the columns you want to check for duplicates (e.g., name, author, narrator, etc.).
   * Go to the Home tab.
   * Click on Remove Duplicates.
4. **Remove Blank Rows**:
   * Go to the Home tab.
   * Click on Remove Rows > Remove Blank Rows.
5. **Handle Non-Numeric Values in the price Column**:
   * Select the price column.
   * Go to the Transform tab.
   * Click on Replace Values.
   * In the Value to Find box, type Free.
   * In the Replace with box, type the number 0.
   * Click OK.
6. **Extract Ratings from stars Column**:
   * Select the stars column.
   * Use Split Column > By Delimiter and choose (other) and enter (star) as the delimiter to split the text into separate columns for stars and ratings.
   * Right-click on the columns with ratings and rename that column (e.g., number of ratings).
   * Remove the ratings text from the ratings column by right-clicking and selecting Replace Values and replace (ratings) with nothing.
   * Do the previous step again to remove the singular (rating).
7. **Remove "out of 5" from the stars Column**:
   * Select the stars column.
   * Go to the Transform tab.
   * Click on Replace Values.
   * In the Value to Find box, type out of 5.
   * Leave the Replace with box empty and click OK.
8. **Remove Rows with Missing Stars or Ratings**:
   * Ratings are crucial to our analysis, so we’ve decided to filter out rows where the stars column or the number of ratings column has Not rated yet or null values.
   * Click the down arrow icon to the right of the number of ratings column header.
   * Uncheck null to remove those rows.
   * Click OK.
9. **Check and Correct Data Types**:
   * Ensure each column has the correct data type by selecting the column and choosing the appropriate data type from the Transform tab.
   * **Change the Data Type for the price Column to Currency**:
     + Select the price column.
     + Go to the Transform tab.
     + Click on the data type dropdown next to Data Type:.
     + Select Currency.
   * **Transform the price Column by Dividing by 100**:
     + Select the price column.
     + Go to the Transform tab.
     + Click on Standard > Divide.
     + Enter 100 and click OK.
   * **Change the Data Type for the number of ratings Column to Whole Number**:
     + Select the number of ratings column.
     + Go to the Transform tab.
     + Click on the data type dropdown next to Data Type:.
     + Select Whole Number.
10. **Check Column Profiles for Outliers**:
    * Go to the View tab.
    * Check the Column Profile option.
    * Review the statistics for each column to see if the range is reasonable.
    * If there are outliers that need to be removed, use the filter options in Power Query to exclude those values.
11. **Load the Cleaned Data into Excel**:
    * Once you have finished cleaning the data, click Close & Load in the Home tab.
    * The cleaned data will be loaded into a new worksheet in Excel.

Remove values from language column so they have to fill up or down values.

Also create duplicates.

Create Blank rows.

Add in some negative prices and stars that don’t make sense, way more than 5.