Contents

[Cleaning the Titanic Dataset in Excel's Power Query Editor 1](#_Toc171906808)

[Step 1: Open Excel and Load the Dataset 1](#_Toc171906809)

[Step 2: Access Power Query 1](#_Toc171906810)

[Step 3: Split the Name Column 1](#_Toc171906811)

[Step 4: Extract Titles from the Title and First Name Column 2](#_Toc171906812)

[Step 5: Create a Conditional Column for Survival 2](#_Toc171906813)

[Step 6: Create a New Passenger Class Column Using Column by Examples 2](#_Toc171906814)

[Step 7: Create a New Family Size Column 3](#_Toc171906815)

[Step 8: Replace Values in the Embarked Column 3](#_Toc171906816)

[Step 9: Check Data Types for All Columns 3](#_Toc171906817)

[Step 10: Remove the Cabin Column 4](#_Toc171906818)

[Step 11: Load the Cleaned Data into Excel 4](#_Toc171906819)

# Cleaning the Titanic Dataset in Excel's Power Query Editor

## Step 1: Open Excel and Load the Dataset

1. Open Microsoft Excel.
2. Go to the File tab, select Open, and choose your Titanic dataset CSV file.

## Step 2: Access Power Query

1. Go to the Data tab on the Excel ribbon.
2. In the Get & Transform Data group, click on Get Data > From File > From Text/CSV.
3. Select your Titanic dataset CSV file and click Import.
4. In the Navigator window, select the sheet or table you want to import and click Transform Data.

## Step 3: Split the Name Column

1. Select the Name column by clicking on the column header.
2. Go to the Home tab.
3. Click on Split Column > By Delimiter.
4. In the Split Column by Delimiter dialog, choose Comma as the delimiter and split at Left-Most delimiter and click OK.
5. Rename the new columns to Last Name and Title and First Name:
   * Right-click on the first new column (Name.1), select Rename, and type Last Name.
   * Right-click on the second new column (Name.2), select Rename, and type Title and First Name.

## Step 4: Extract Titles from the Title and First Name Column

1. Select the Title and First Name column.
2. Go to the Transform tab.
3. Click on Split Column > By Delimiter.
4. In the Split Column by Delimiter dialog, choose Custom as the delimiter and enter . in the custom delimiter field. Choose split at Left-Most delimiter and click OK.
5. Rename the new columns to Title and First Name:
   * Right-click on the first new column (Name.1), select Rename, and type Title.
   * Right-click on the second new column (Name.2), select Rename, and type First Name.

## Step 5: Create a Conditional Column for Survival

1. Go to the Add Column tab.
2. Click on Conditional Column.
3. In the New column name box, type Survived\_YN.
4. Set the condition to If [Survived] equals 1 then "Yes" else "No".
5. Click OK.
6. Check the data type:
   * Select the Survived\_YN column.
   * Go to the Transform tab.
   * Ensure the data type is Text.
7. Remove the original Survived column:
   * Select the Survived column.
   * Right-click and choose Remove.

## Step 6: Create a New Passenger Class Column Using Column by Examples

1. Select the Pclass column.
2. Go to the Add Column tab.
3. Click on Column From Examples > From Selection.
4. In the rows of the new column, type:
   * When the value in the Pclass column is 1, type First Class.
   * For 2, type Second Class.
   * For 3, type Third Class.
5. Power Query will auto-fill the rest. Click OK.
6. Rename the new column to Passenger Class.
7. Remove the original Pclass column:
   * Select the Pclass column.
   * Right-click and choose Remove.
8. Check the data type:
   * Select the Passenger Class column.
   * Go to the Transform tab.
   * Ensure the data type is Text.

## Step 7: Create a New Family Size Column

1. Go to the Add Column tab.
2. Click on Custom Column.
3. In the New column name box, type Family Size.
4. In the Custom column formula box, type =[SibSp] + [Parch] + 1.
5. Click OK.
6. Check the data type:
   * Select the Family Size column.
   * Go to the Transform tab.
   * Ensure the data type is Whole Number.
7. Remove the SibSp and Parch columns:
   * Select the SibSp column.
   * Right-click and choose Remove.
   * Repeat for the Parch column.

## Step 8: Replace Values in the Embarked Column

1. Select the Embarked column.
2. Go to the Transform tab.
3. Click on Replace Values.
4. In the Value to Find box, type C.
5. In the Replace with box, type Cherbourg.
6. Click OK.
7. Repeat the above steps for:
   * S with Southampton
   * Q with Queenstown

## Step 9: Check Data Types for All Columns

1. Ensure each column has the correct data type by selecting the column and choosing the appropriate data type from the Transform tab:
   * PassengerId: Whole Number
   * Last Name: Text
   * Title: Text
   * First Name: Text
   * Sex: Text
   * Age: Decimal Number
   * Ticket: Text
   * Fare: Currency
   * Cabin: Text
   * Embarked: Text
   * Survived\_YN: Text
   * Passenger Class: Text
   * Family Size: Whole Number

## Step 10: Remove the Cabin Column

1. Select the Cabin column.
2. Right-click and choose Remove.

## Step 11: Load the Cleaned Data into Excel

1. Once you have finished cleaning the data, go to the Home tab.
2. Click Close & Load.
3. The cleaned data will be loaded into a new worksheet in Excel.