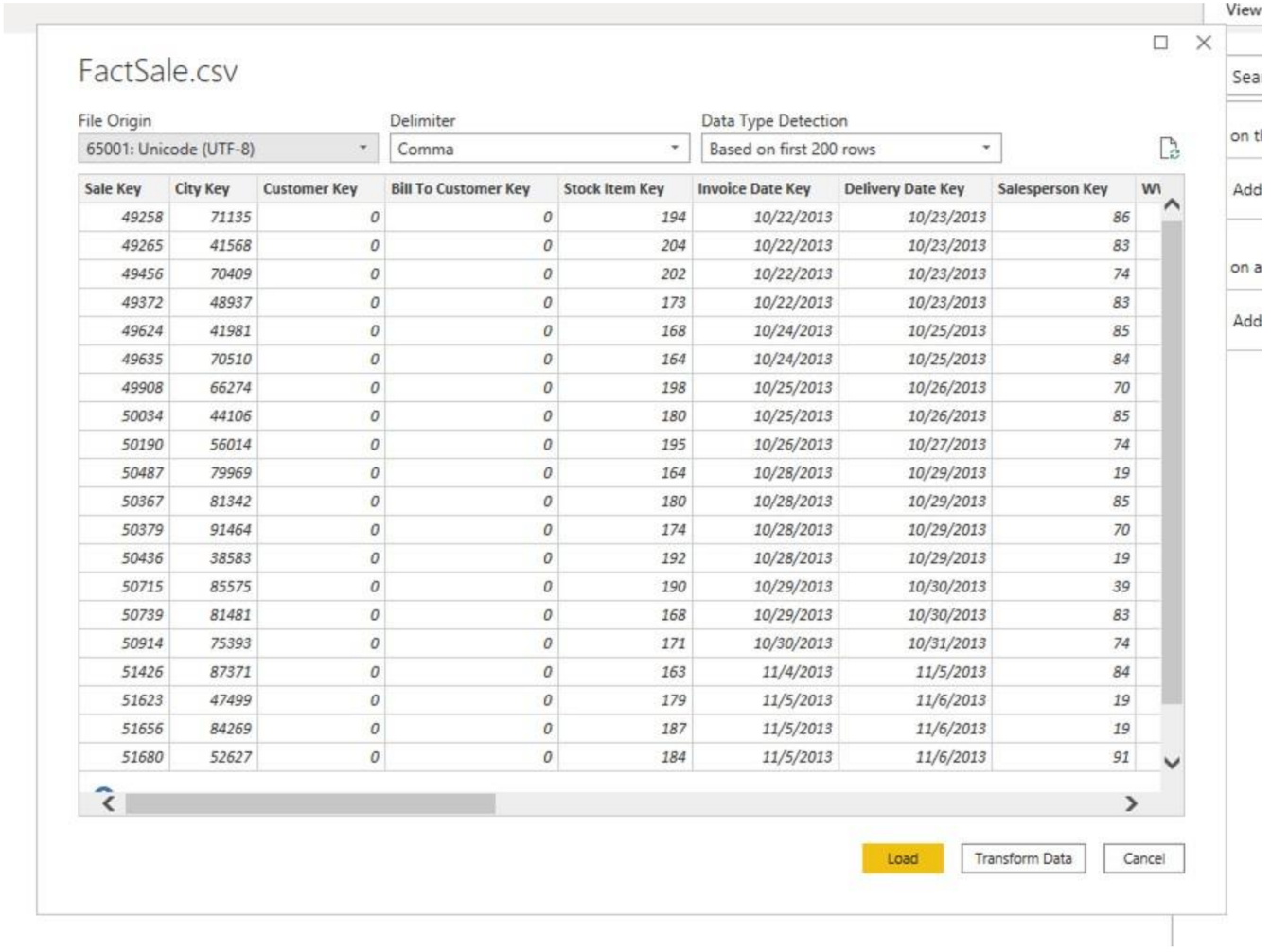
Transforming data

INTRODUCTION TO POWER BI

Transforming data

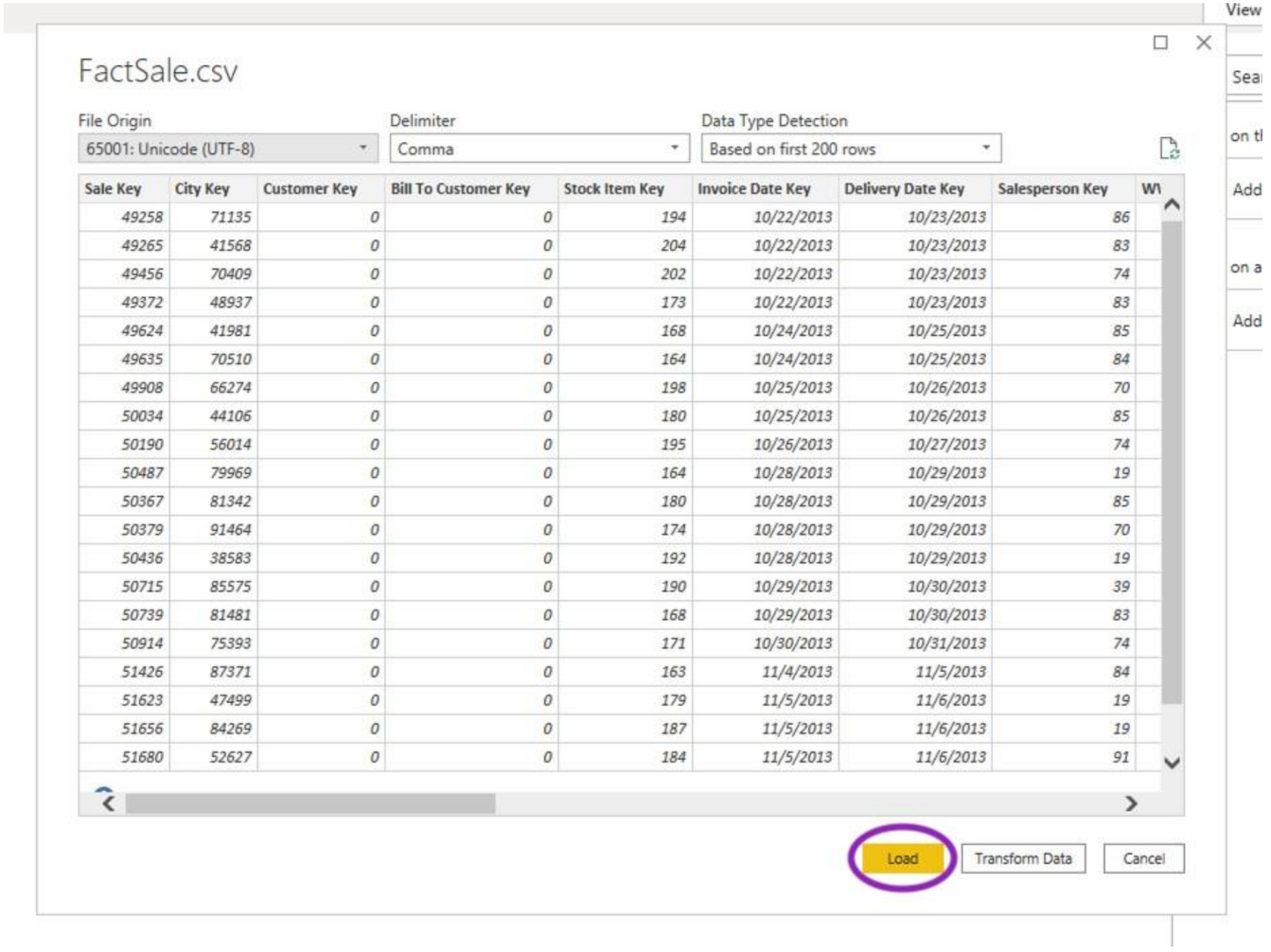
- Dataset may contain:
  - Columns you don't need
  - Inconvenient and inconsistent formatting
  - Extra characters
  - Blank rows
- Cleaning data

Loading data



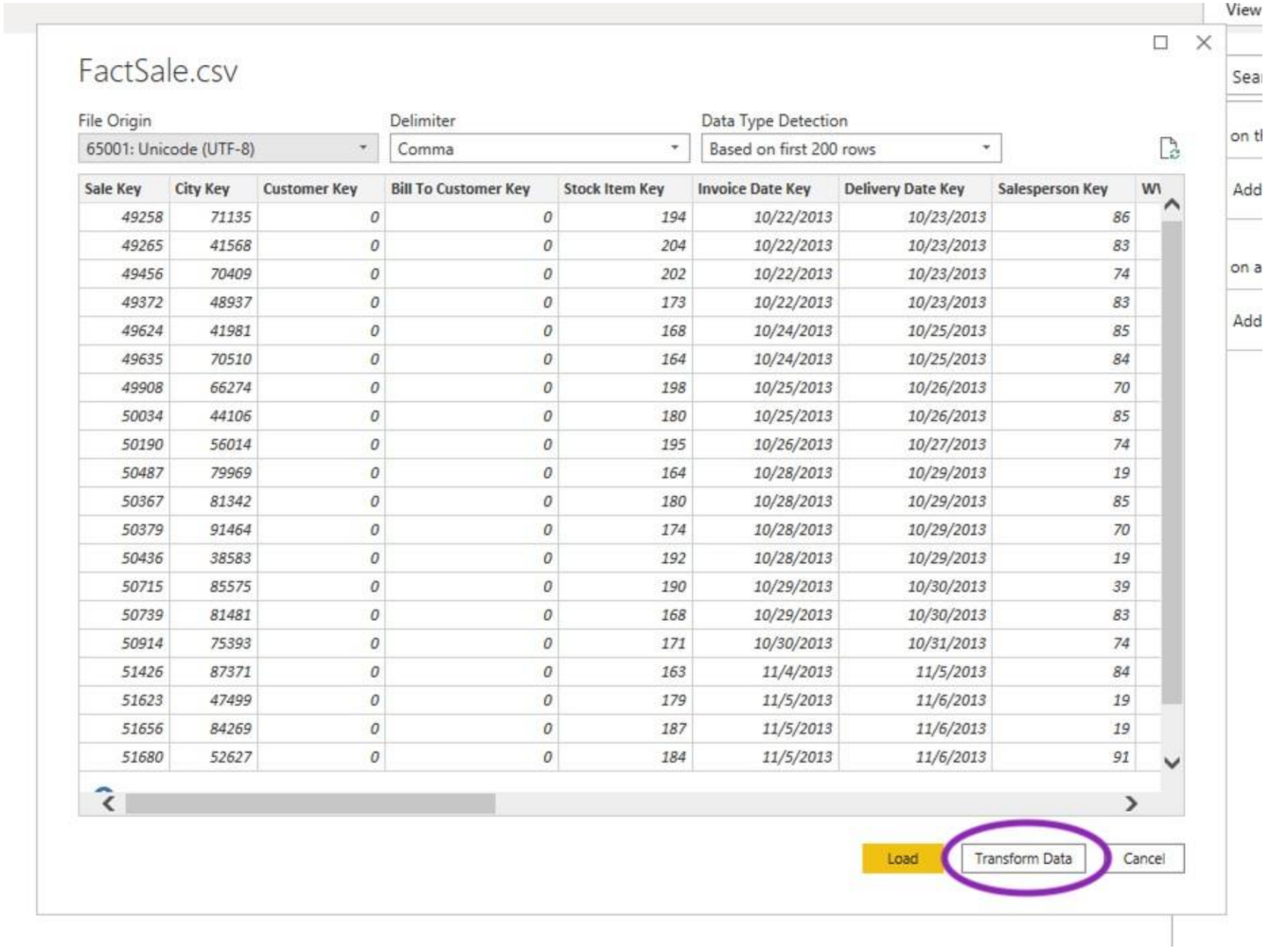
Sale Key	City Key	Customer Key	Bill To Customer Key	Stock Item Key	Invoice Date Key	Delivery Date Key	Salesperson Key
49258	71135	0	0	194	10/22/2013	10/23/2013	86
49265	41568	0	0	204	10/22/2013	10/23/2013	83
49456	70409	0	0	202	10/22/2013	10/23/2013	74
49372	48937	0	0	173	10/22/2013	10/23/2013	83
49624	41981	0	0	168	10/24/2013	10/25/2013	85
49635	70510	0	0	164	10/24/2013	10/25/2013	84
49908	66274	0	0	198	10/25/2013	10/26/2013	70
50034	44106	0	0	180	10/25/2013	10/26/2013	85
50190	56014	0	0	195	10/26/2013	10/27/2013	74
50487	79969	0	0	164	10/28/2013	10/29/2013	19
50367	81342	0	0	180	10/28/2013	10/29/2013	85
50379	91464	0	0	174	10/28/2013	10/29/2013	70
50436	38583	0	0	192	10/28/2013	10/29/2013	19
50715	85575	0	0	190	10/29/2013	10/30/2013	39
50739	81481	0	0	168	10/29/2013	10/30/2013	83
50914	75393	0	0	171	10/30/2013	10/31/2013	74
51426	87371	0	0	163	11/4/2013	11/5/2013	84
51623	47499	0	0	179	11/5/2013	11/6/2013	19
51656	84269	0	0	187	11/5/2013	11/6/2013	19
51680	52627	0	0	184	11/5/2013	11/6/2013	91

Loading data



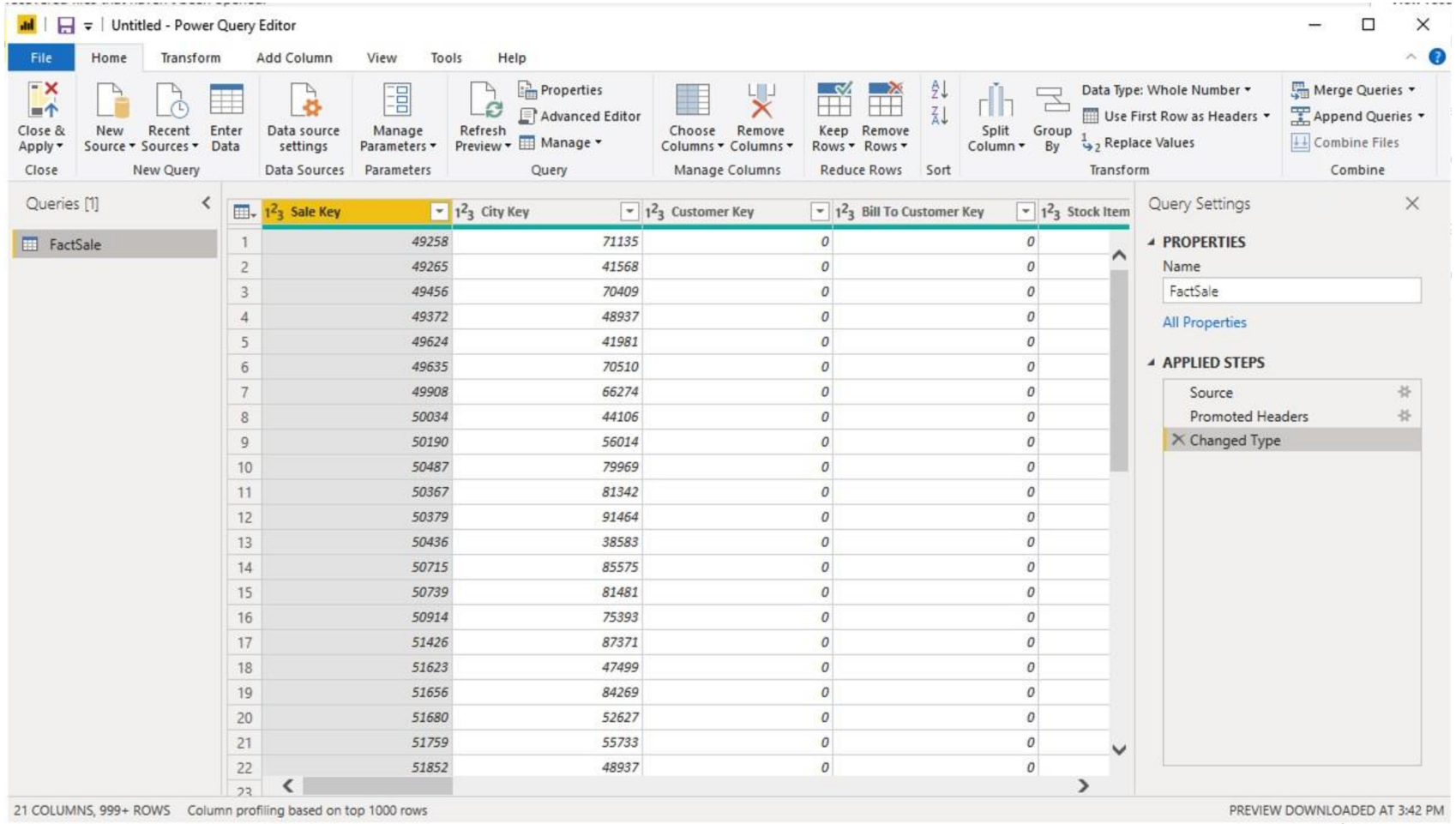
Sale Key	City Key	Customer Key	Bill To Customer Key	Stock Item Key	Invoice Date Key	Delivery Date Key	Salesperson Key	WI
49258	71135	0	0	194	10/22/2013	10/23/2013	86	
49265	41568	0	0	204	10/22/2013	10/23/2013	83	
49456	70409	0	0	202	10/22/2013	10/23/2013	74	
49372	48937	0	0	173	10/22/2013	10/23/2013	83	
49624	41981	0	0	168	10/24/2013	10/25/2013	85	
49635	70510	0	0	164	10/24/2013	10/25/2013	84	
49908	66274	0	0	198	10/25/2013	10/26/2013	70	
50034	44106	0	0	180	10/25/2013	10/26/2013	85	
50190	56014	0	0	195	10/26/2013	10/27/2013	74	
50487	79969	0	0	164	10/28/2013	10/29/2013	19	
50367	81342	0	0	180	10/28/2013	10/29/2013	85	
50379	91464	0	0	174	10/28/2013	10/29/2013	70	
50436	38583	0	0	192	10/28/2013	10/29/2013	19	
50715	85575	0	0	190	10/29/2013	10/30/2013	39	
50739	81481	0	0	168	10/29/2013	10/30/2013	83	
50914	75393	0	0	171	10/30/2013	10/31/2013	74	
51426	87371	0	0	163	11/4/2013	11/5/2013	84	
51623	47499	0	0	179	11/5/2013	11/6/2013	19	
51656	84269	0	0	187	11/5/2013	11/6/2013	19	
51680	52627	0	0	184	11/5/2013	11/6/2013	91	

Loading data

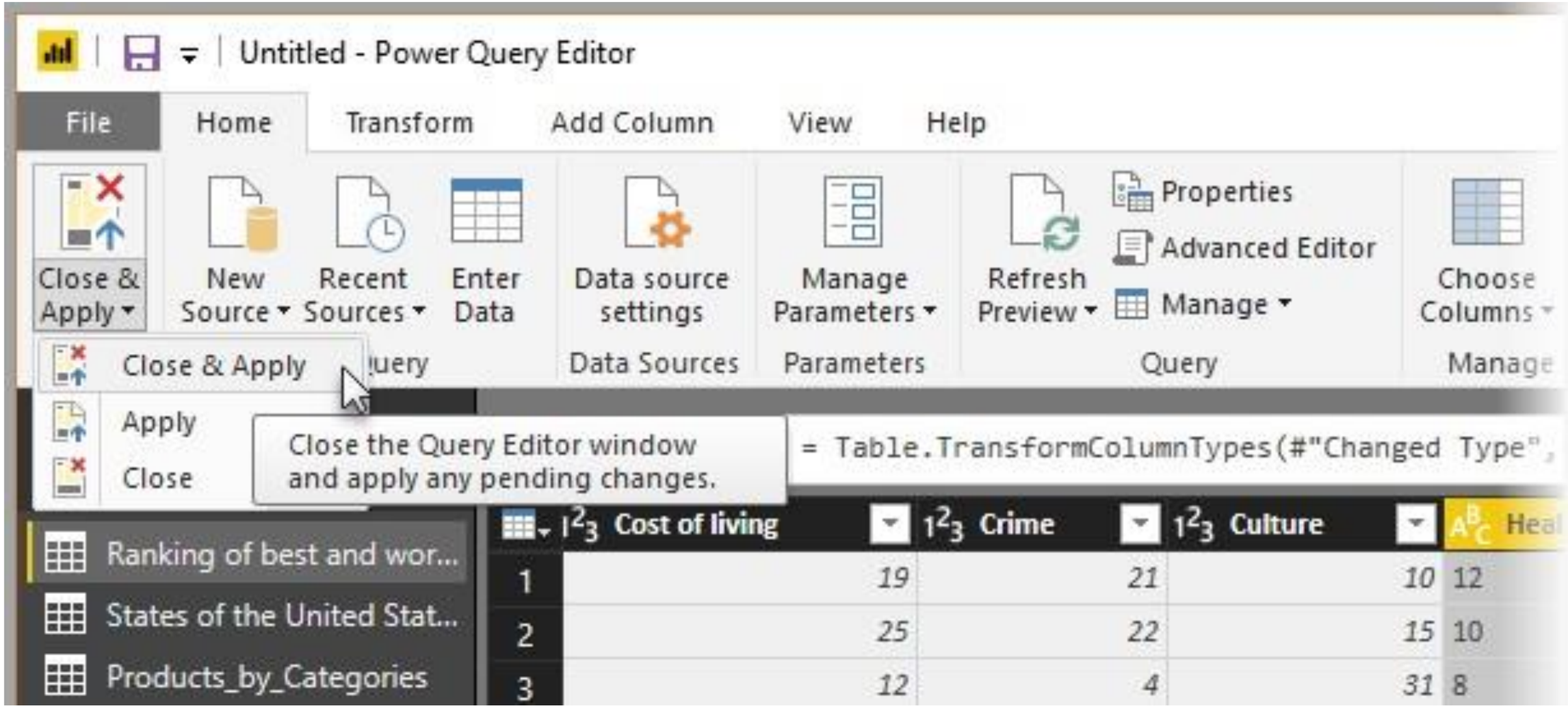


Sale Key	City Key	Customer Key	Bill To Customer Key	Stock Item Key	Invoice Date Key	Delivery Date Key	Salesperson Key
49258	71135	0	0	194	10/22/2013	10/23/2013	86
49265	41568	0	0	204	10/22/2013	10/23/2013	83
49456	70409	0	0	202	10/22/2013	10/23/2013	74
49372	48937	0	0	173	10/22/2013	10/23/2013	83
49624	41981	0	0	168	10/24/2013	10/25/2013	85
49635	70510	0	0	164	10/24/2013	10/25/2013	84
49908	66274	0	0	198	10/25/2013	10/26/2013	70
50034	44106	0	0	180	10/25/2013	10/26/2013	85
50190	56014	0	0	195	10/26/2013	10/27/2013	74
50487	79969	0	0	164	10/28/2013	10/29/2013	19
50367	81342	0	0	180	10/28/2013	10/29/2013	85
50379	91464	0	0	174	10/28/2013	10/29/2013	70
50436	38583	0	0	192	10/28/2013	10/29/2013	19
50715	85575	0	0	190	10/29/2013	10/30/2013	39
50739	81481	0	0	168	10/29/2013	10/30/2013	83
50914	75393	0	0	171	10/30/2013	10/31/2013	74
51426	87371	0	0	163	11/4/2013	11/5/2013	84
51623	47499	0	0	179	11/5/2013	11/6/2013	19
51656	84269	0	0	187	11/5/2013	11/6/2013	19
51680	52627	0	0	184	11/5/2013	11/6/2013	91

Power Query Editor



Don't forget to Close & Apply



Let's practice!

INTRODUCTION TO POWER BI

Power Query Editor

INTRODUCTION TO POWER BI

Raw data usually doesn't arrive in the perfect form when you account for things like human errors, bugs, and file conversion. Power BI accounts for this with the Power Query Editor which allows you to transform data before loading it. In this exercise, you will load another dimension called `DimCustomer`, except unlike the others, this file will need to be edited prior to loading.

- Open the **csv** file `DimCustomer.csv` from `Datasets/WWI` folder on the desktop.
- Select the *Transform Data* button

- Remove the first row. It contains mostly blanks and does not provide any information.
- Make the resulting first row the header row.

- Delete the columns `Valid From` and `Valid To`.
- Close and apply.

In the *Fields* pane, select the *Edit Query* menu option from `DimCustomer`.

In Power Query Editor window that has opened, how many steps are listed in the "Applied Steps" section?

Now that we've loaded `DimCustomer` and cleaned it up, let's create a visualization with it. Power BI automatically **sums** up numerical fields in a visualization. For example, the default is to show the sum of all the sale totals, rather than the average sale total. In this exercise, you will try a new aggregation.

- Click "Close & Apply" to close Power Query in case it is still open.
- Check that Power BI recognized a relationship between `FactSale` and `DimCustomer`.
- Add a new page tab in the *Report* view.

Make a Clustered Column Chart using `Buying Group` from `DimCustomer` and `Total Including Tax` from `FactSale`.

Change it so that the value is the **minimum** of `Total Including Tax`.

According to total Including tax, how much was the cheapest sale made to Tailspin Toys (answer format: 1.1)?

Transformations

Earlier you practiced cleaning data at row-level, like deleting erroneous rows or changing the header row. Now, we'll take a look at issues at the column-level.

- Make sure no bars are selected on the bar graph.
- Create a Card visualization with the value `Credit Limit` from `DimCustomer`.

The card should show **?** -, which is unexpected! Edit the query of `DimCustomer` to open up the Power Query Editor and fix the `Credit Limit` column.

- Replace values so that `?` s are replaced with blanks in `Credit Limit`.
- Repeat so that `-` s are replaced with blanks for the `Credit Limit` column.

Change the data type of `Credit Limit` from *Text* to *Decimal Number*.

Close and apply and return to the *Report* view. In the card, change the value to be the **average** `Credit Limit`.

What is the average credit card limit of a Wide World Importers customer (answer format: $11.11K)?

Now that you know more about formatting data types. Let's go back to the first report you made, which should be the first page tab titled "Sales Data". Take a look at the formatting of the data in the table. Let's improve the formatting of the `Profit` and the `Total Including Tax` columns so it's immediately clear they are monetary values, unlike `Quantity`.

- In the *Data* view of `FactSale`, select the `Total Including Tax` column.
- Using *Column tools*, change the format to *Currency*.

Change the number of decimal places shown to *2* instead of *Auto*.

- Change the default aggregation from *Sum* to *Average*.
- Repeat the same format and decimal place changes to the `Profit` column.

- Add a card to your report and select `Total Including Tax`.
- If applicable, clear any selections on the `Employee` slicer so that all employees are considered.

What does the "Total Including Tax" card now display? (answer format: $111.11)

Maps are an engaging way to present data with a geographic layer. Imagine we wanted to depict the profit each state in the US generates. We could create a bar chart showing the states and the profit they generate. However, since there are 50 states, a map is much easier to scan for patterns and outliers.

- Load the dimension table `DimCity.csv` from the `Datasets/WWI` folder on the Desktop.
- Go to the *Model* view and make sure a relationship is found between `DimCity` and `FactSale`.

In the *Data* view, change the *Data category* of `DimCity`'s `State Province` to "State or Province".

Make sure the default summarization for `Profit` from `FactSale` is "Average".

- In the *Report* view, navigate to the second tab.
- Create a *Map* visualization using `State Province` as *Location* and `Profit` as *Bubble size*.

Add a *Slicer* for the `Buying Group` field from the `DimCustomer` table. Arrange the report to your liking and add a title.

Using the map and the slicer, which state generates the highest average profit for the "Wingtip Toys"?

- ○ Washington
- ○ Oregon
- ○ Alaska
- ○ California