Data Visualization Tools and Software

Aim : Prepare Data in Power BI Desktop : Connect and transform Raw data

Task:

1. Install Power-BI

2. Prepare data in Power-BI

In this lab, we learn how to:

Open Power BI Desktop
Set Power BI Desktop options
Types of Data Connectors;

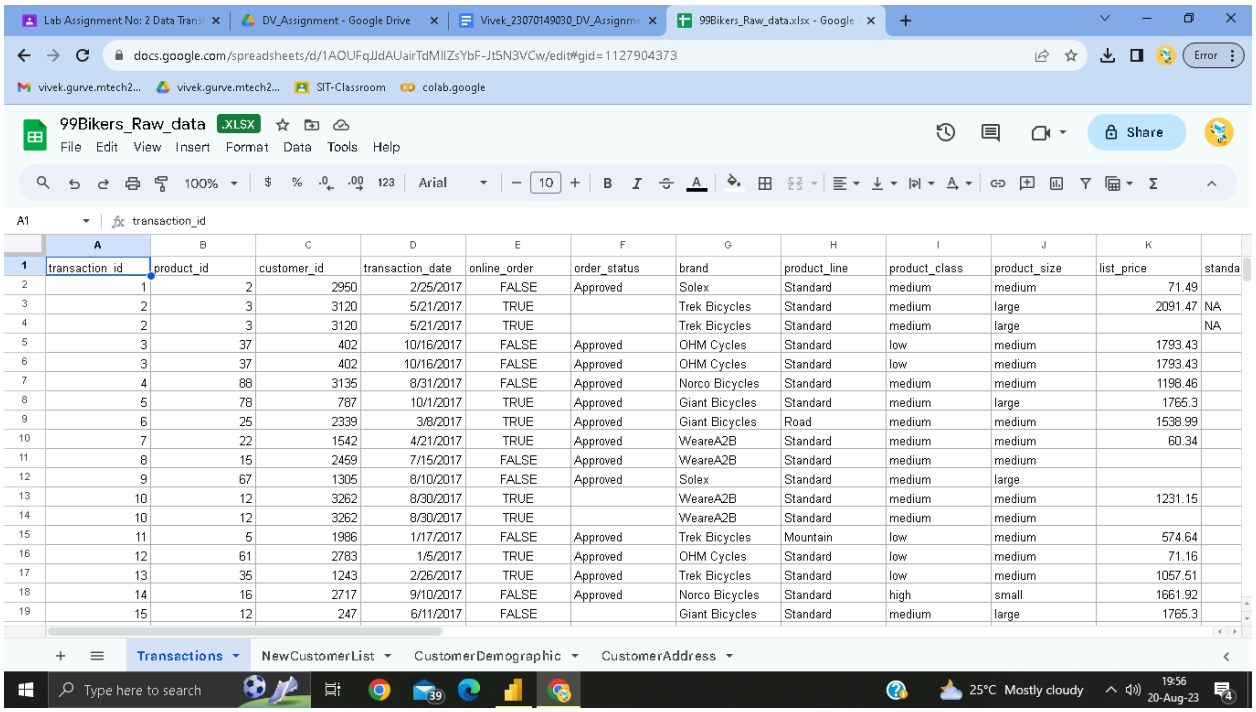
- Connect to source data (Excel/CSV/etc.)
- Load data

Preview source data
Explore Transform Data options

- Clean data
- Remove/Replace Null values, duplicate values, missing value
- Merge / split columns
- Add columns, remove columns
- Change data types

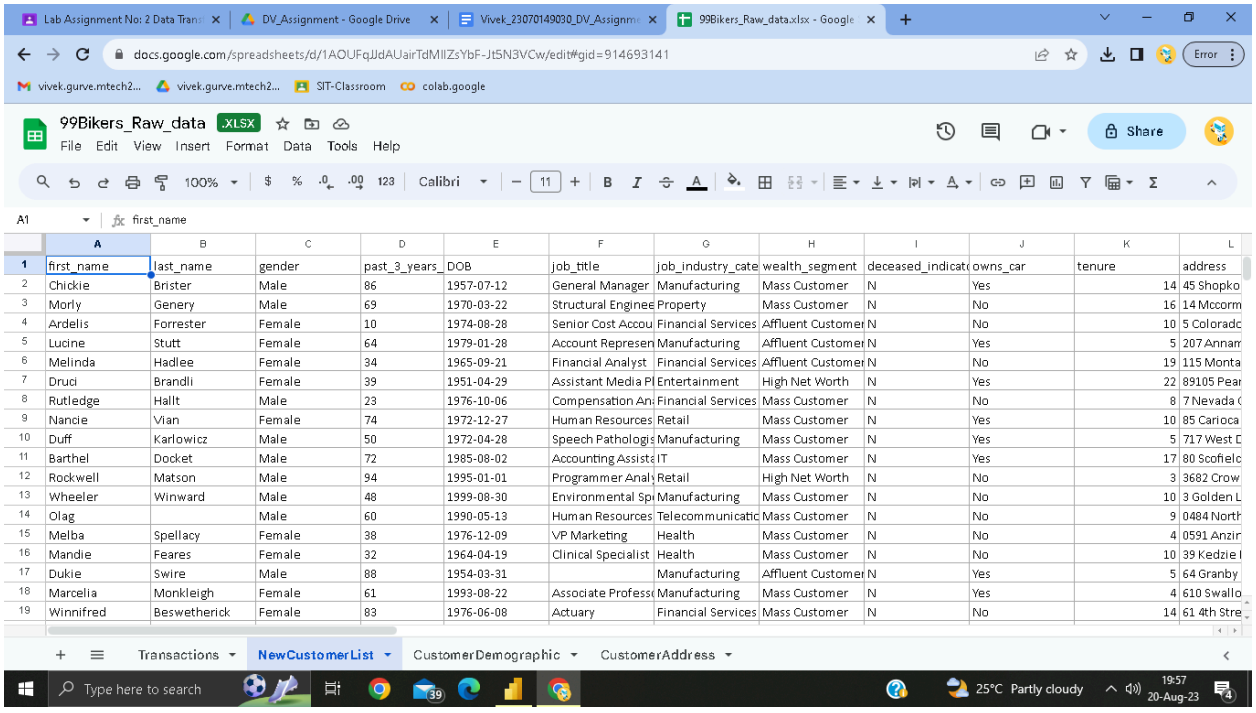
Preview of Data

Dataset named as "**99Bikers_Raw_data**"



transaction_id	product_id	customer_id	transaction_date	online_order	order_status	brand	product_line	product_class	product_size	list_price	standa
1	2	2950	2/25/2017	FALSE	Approved	Solex	Standard	medium	medium	71.49	
2	3	3120	5/21/2017	TRUE		Trek Bicycles	Standard	medium	large	2091.47	NA
2	3	3120	5/21/2017	TRUE		Trek Bicycles	Standard	medium	large		NA
3	37	402	10/16/2017	FALSE	Approved	OHM Cycles	Standard	low	medium	1793.43	
3	37	402	10/16/2017	FALSE	Approved	OHM Cycles	Standard	low	medium	1793.43	
4	88	3135	8/31/2017	FALSE	Approved	Norco Bicycles	Standard	medium	medium	1198.46	
5	78	787	10/1/2017	TRUE	Approved	Giant Bicycles	Standard	medium	large	1765.3	
6	25	2339	3/8/2017	TRUE	Approved	Giant Bicycles	Road	medium	medium	1538.99	
7	22	1542	4/21/2017	TRUE	Approved	WeareA2B	Standard	medium	medium	60.34	
8	15	2459	7/15/2017	FALSE	Approved	WeareA2B	Standard	medium	medium		
9	67	1305	8/10/2017	FALSE	Approved	Solex	Standard	medium	large		
10	12	3262	8/30/2017	TRUE		WeareA2B	Standard	medium	medium	1231.15	
10	12	3262	8/30/2017	TRUE		WeareA2B	Standard	medium	medium		
11	5	1986	1/17/2017	FALSE	Approved	Trek Bicycles	Mountain	low	medium	574.64	
12	61	2783	1/5/2017	TRUE	Approved	OHM Cycles	Standard	low	medium	71.16	
13	35	1243	2/26/2017	TRUE	Approved	Trek Bicycles	Standard	low	medium	1057.51	
14	16	2717	9/10/2017	FALSE	Approved	Norco Bicycles	Standard	high	small	1661.92	
15	12	247	6/11/2017	FALSE		Giant Bicycles	Standard	medium	large	1765.3	

Fig.-Transactions



first_name	last_name	gender	past_3_years_	DOB	job_title	job_industry_cate	wealth_segment	deceased_indicat	owns_car	tenure	address
Chickie	Brister	Male	86	1957-07-12	General Manager	Manufacturing	Mass Customer	N	Yes	14	45 Shopko
Morly	Genery	Male	69	1970-03-22	Structural Enginee	Property	Mass Customer	N	No	16	14 Mccorm
Ardelis	Forrester	Female	10	1974-08-28	Senior Cost Accou	Financial Services	Affluent Customer	N	No	10	5 Colorado
Lucine	Stutt	Female	64	1979-01-28	Account Represen	Manufacturing	Affluent Customer	N	Yes	5	207 Annam
Melinda	Hadlee	Female	34	1965-09-21	Financial Analyst	Financial Services	Affluent Customer	N	No	19	115 Monta
Druci	Brandli	Female	39	1951-04-29	Assistant Media Pl	Entertainment	High Net Worth	N	Yes	22	89105 Pear
Rutledge	Hallt	Male	23	1976-10-06	Compensation Ani	Financial Services	Mass Customer	N	No	8	7 Nevada C
Nancie	Vian	Female	74	1972-12-27	Human Resources	Retail	Mass Customer	N	Yes	10	85 Carioca
Duff	Karlowicz	Male	50	1972-04-28	Speech Pathologis	Manufacturing	Mass Customer	N	Yes	5	717 West D
Barthel	Docket	Male	72	1985-08-02	Accounting Assista	IT	Mass Customer	N	Yes	17	80 Scofield
Rockwell	Matson	Male	94	1995-01-01	Programmer Analy	Retail	High Net Worth	N	No	3	3682 Crow
Wheeler	Winward	Male	48	1999-08-30	Environmental Spe	Manufacturing	Mass Customer	N	No	10	3 Golden L
Olag		Male	60	1990-05-13	Human Resources	Telecommunicatio	Mass Customer	N	No	9	0484 North
Melba	Spellacy	Female	38	1976-12-09	VP Marketing	Health	Mass Customer	N	No	4	0591 Anzin
Mandie	Feares	Female	32	1964-04-19	Clinical Specialist	Health	Mass Customer	N	No	10	39 Kedzie I
Dukie	Swire	Male	88	1954-03-31		Manufacturing	Affluent Customer	N	Yes	5	64 Granby
Marcelia	Monkleigh	Female	61	1993-08-22	Associate Professo	Manufacturing	Mass Customer	N	Yes	4	610 Swallo
Winnifred	Beswetherick	Female	83	1976-06-08	Actuary	Financial Services	Mass Customer	N	No	14	61 4th Stre

Fig.- NewCustomerList

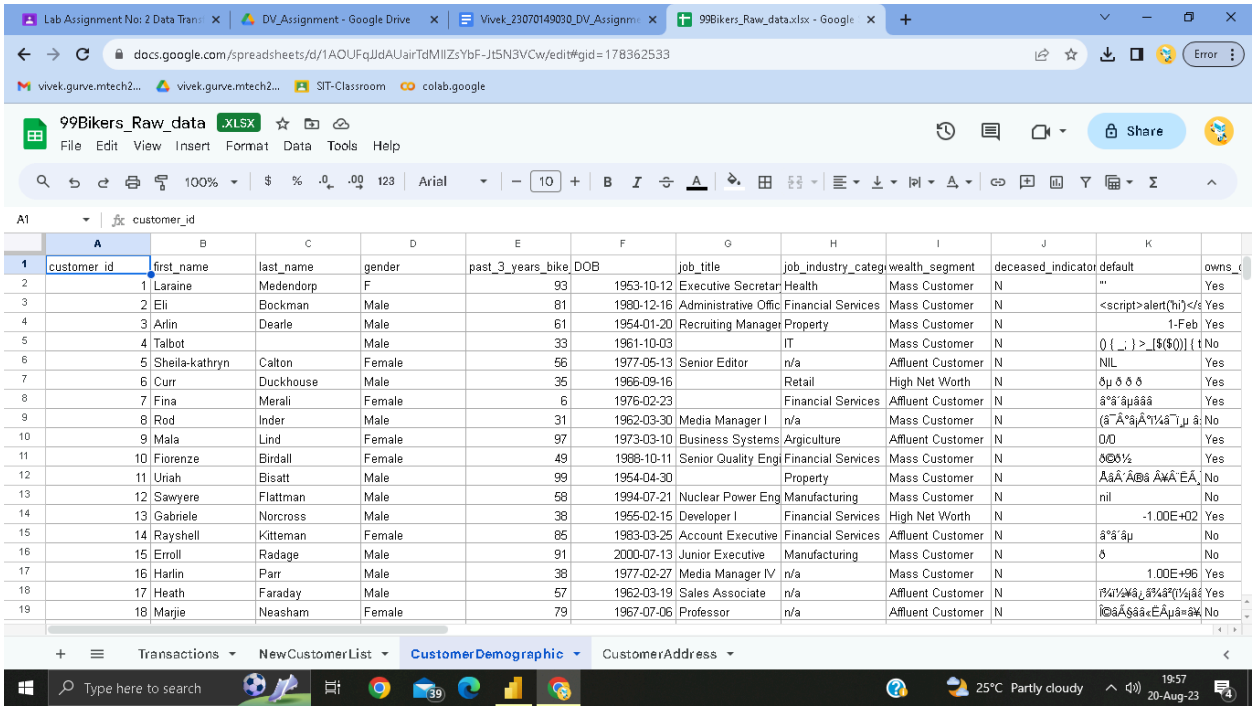


Fig.- CustomerDemographic

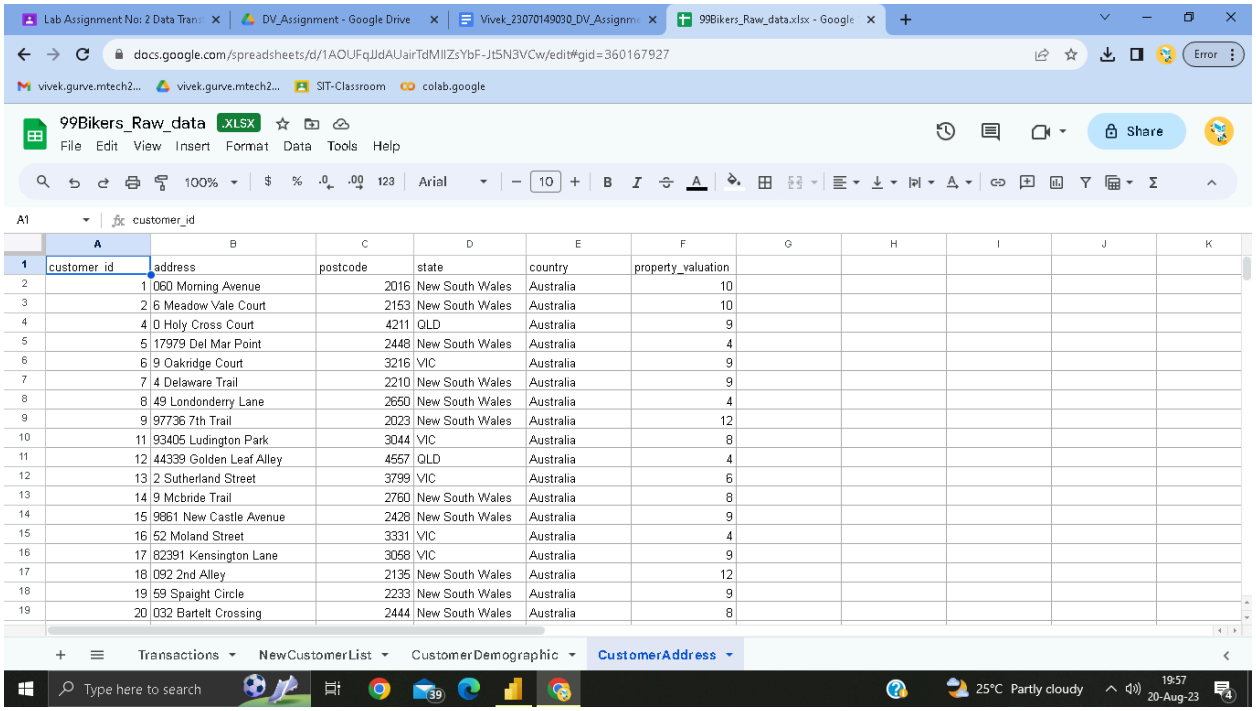


Fig.- CustomerAddress

Data Transformation

Power BI Power Query Editor can be used to edit or transform data files before they get loaded into the Power BI dashboard. The Query Editor serves as an intermediate data container that allows you to modify data by choosing columns and rows, pivoting and unpivoting columns, splitting columns and rows, etc.

Power BI Desktop has three views:

- Report view – use queries that you create to build compelling visualizations, arranged as we want them to appear, and with multiple pages, that you can share with others.
- Data view – See the data in your report in data model format, where we can add measures, create new columns, and manage relationships.
- Model view – Get a graphical representation of the relationships that are established in your data model, and manage or modify them as needed.

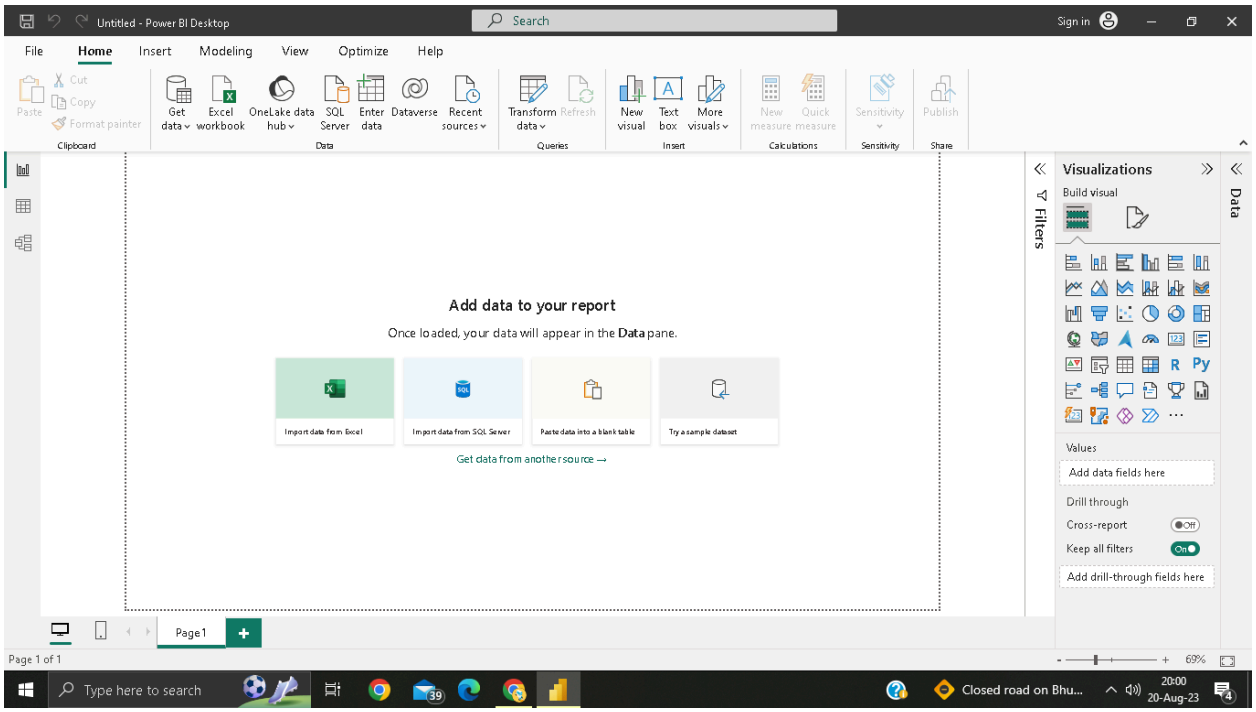


Fig.- Power BI Desktop

Get Data

Select the Text/CSV option in Get Data. This action launches a local file browser where you can select your text file. Select Open to open the file. From the Navigator, you can either transform the data in the Power Query Editor by selecting Transform Data, or load the data by selecting Load.

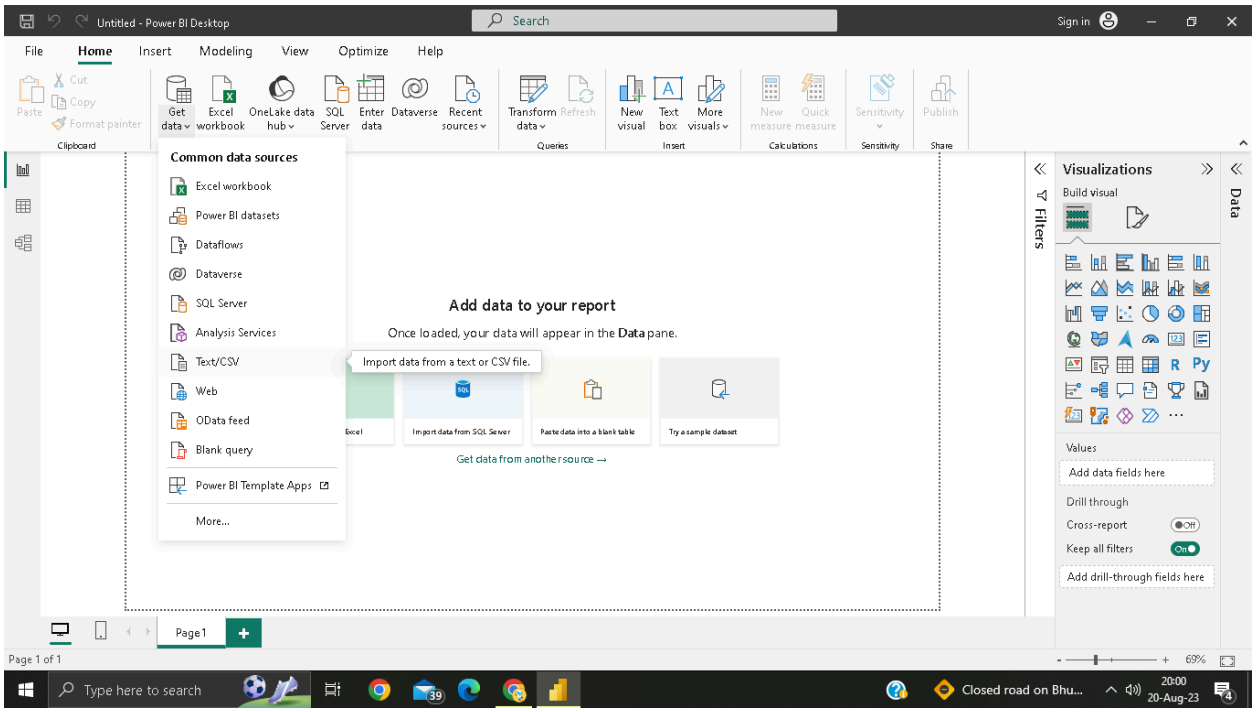


Fig.- Importing data from a CSV file

Navigator

The Navigator window displays a preview. At this point, you can edit the query before loading the table, by selecting Transform Data from the bottom of the window, or just load the table.

From the Navigator, select Transform Data to begin transforming the data in the Power Query Editor.

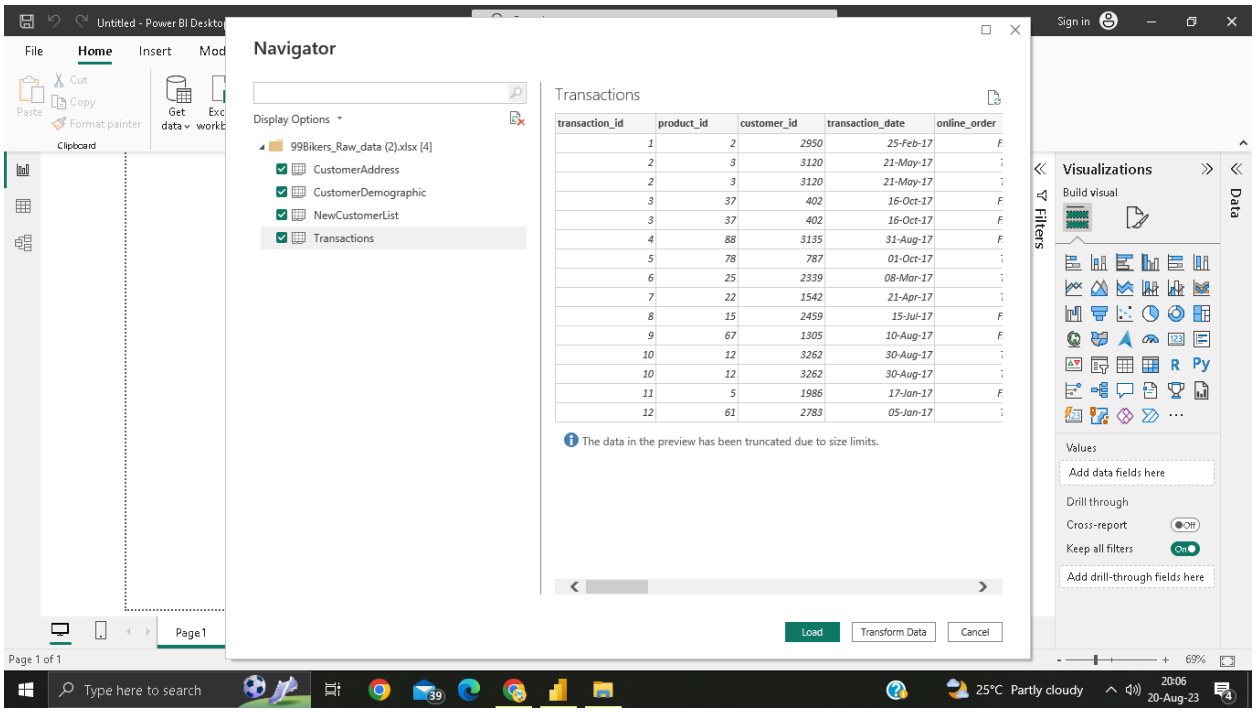


Fig.- Navigator from a CSV file

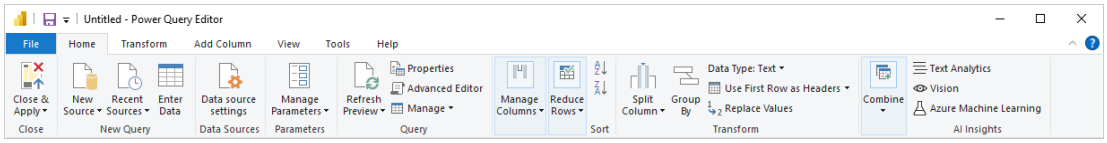
Power Query Editor

Use Power Query Editor to connect to one or many data sources, shape and transform the data to meet our needs, then load that model into Power BI Desktop.

The query ribbon

The ribbon in Power Query Editor consists of four tabs: Home, Transform, Add Column, View, Tools, and Help.

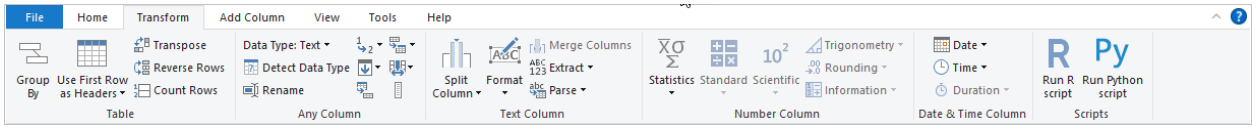
The Home tab contains the common query tasks.



Transform tab

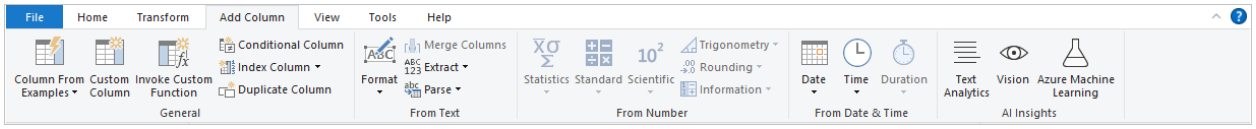
The Transform tab provides access to common data transformation tasks, such as:

- Adding or removing columns
- Changing data types
- Splitting columns
- Other data-driven tasks



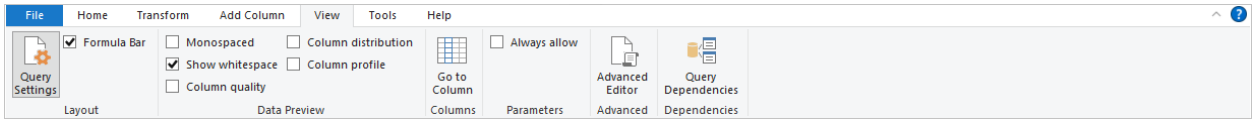
Add Column

The Add Column tab provides more tasks associated with adding a column, formatting column data, and adding custom columns. The following image shows the Add Column tab.



View tab

The View tab on the ribbon is used to toggle whether certain panes or windows are displayed. It's also used to display the Advanced Editor. The following image shows the View tab.



Extract

Creating a new column containing a specified number of characters from the start of each value in the selected column.

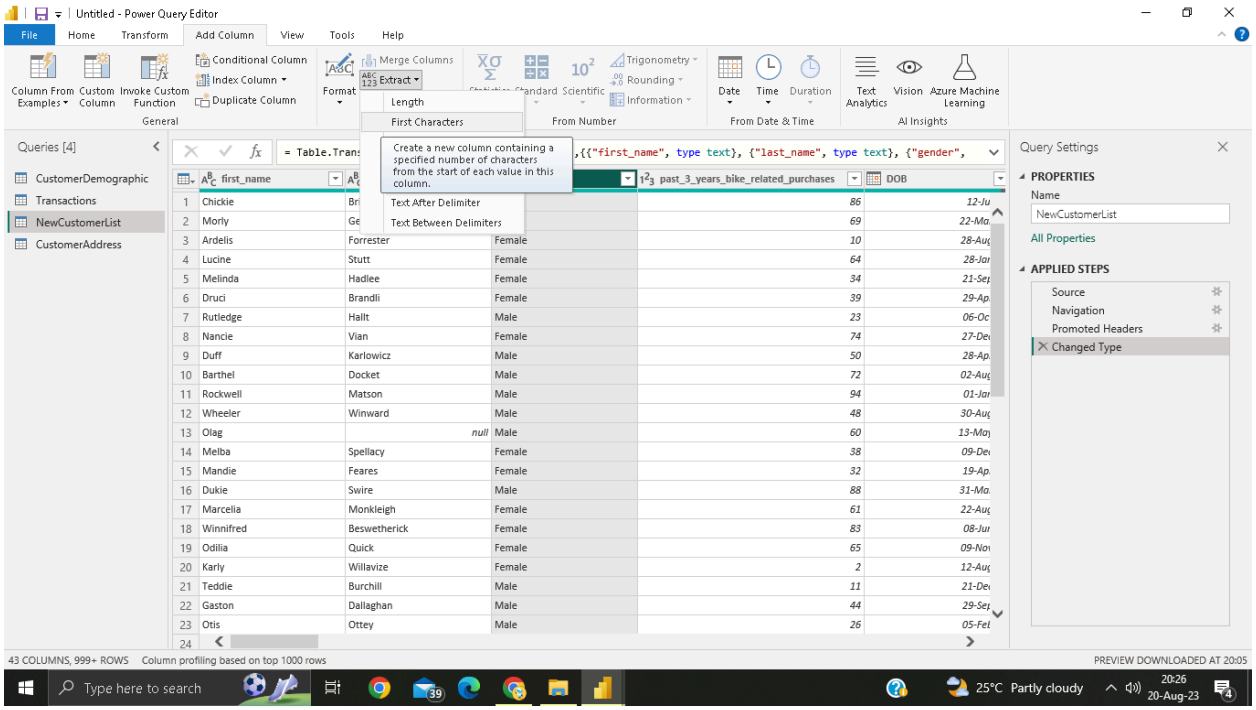


Fig.-Extract first character from gender column

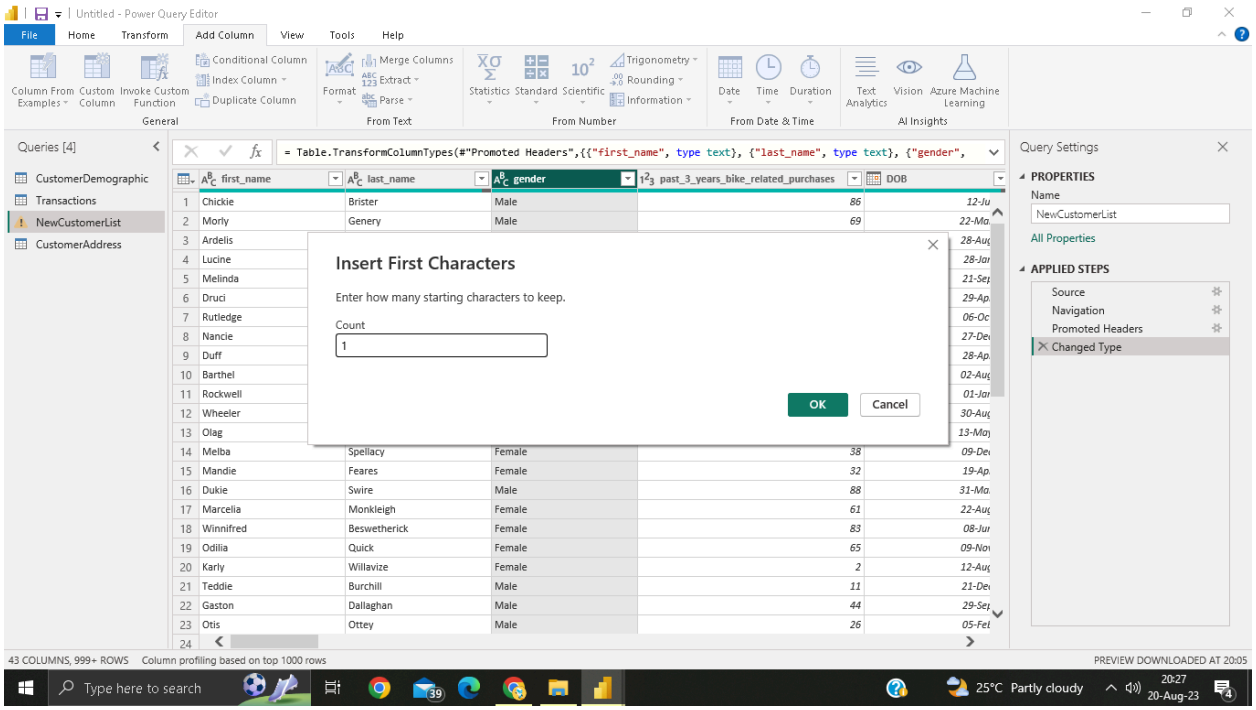


Fig.-Enter how many starting character to keep

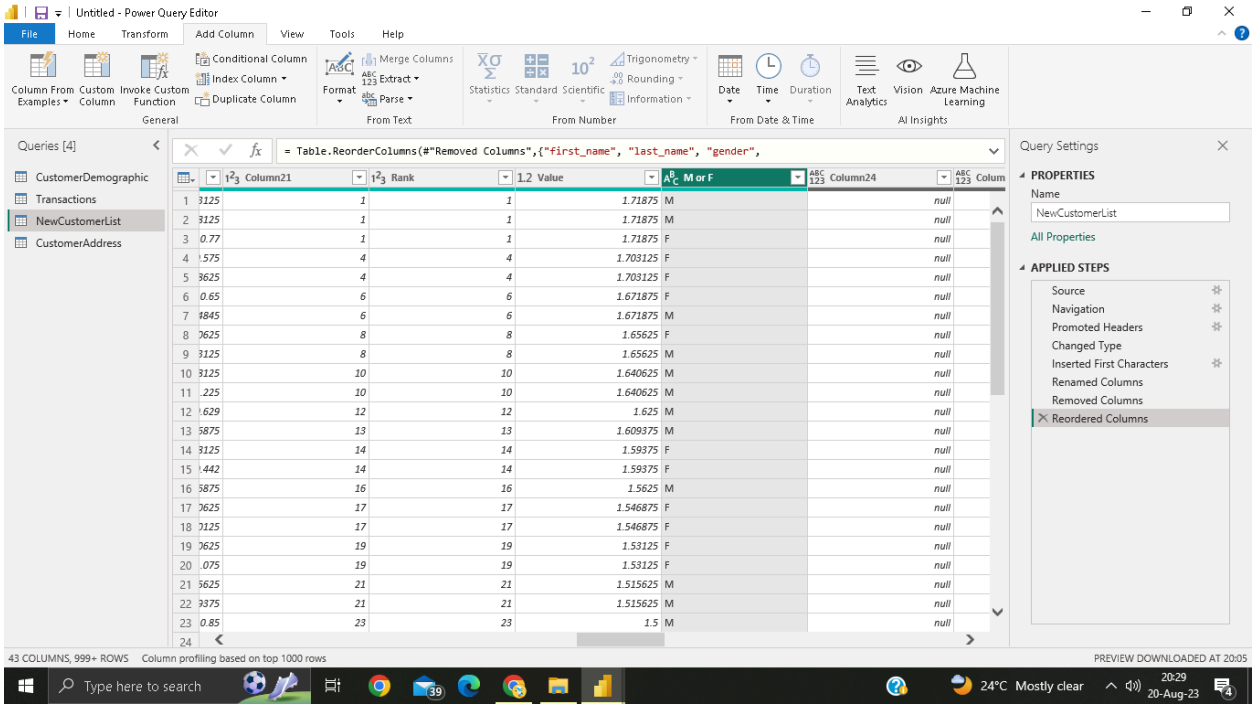


Fig.-M or F column extracted from gender column

Change Data Type

Converts the values from the Any data type to a data type based on the inspection of the values from each column.

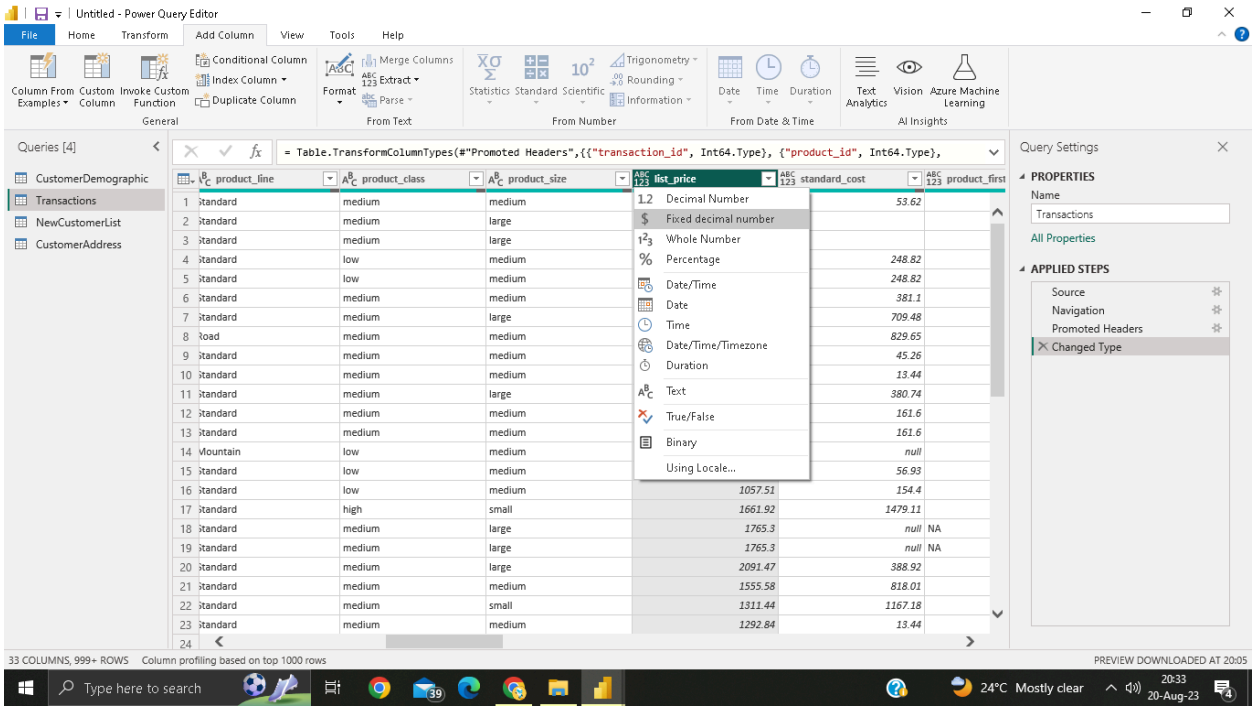


Fig.-Change list_price Data Type to Fixed decimal number

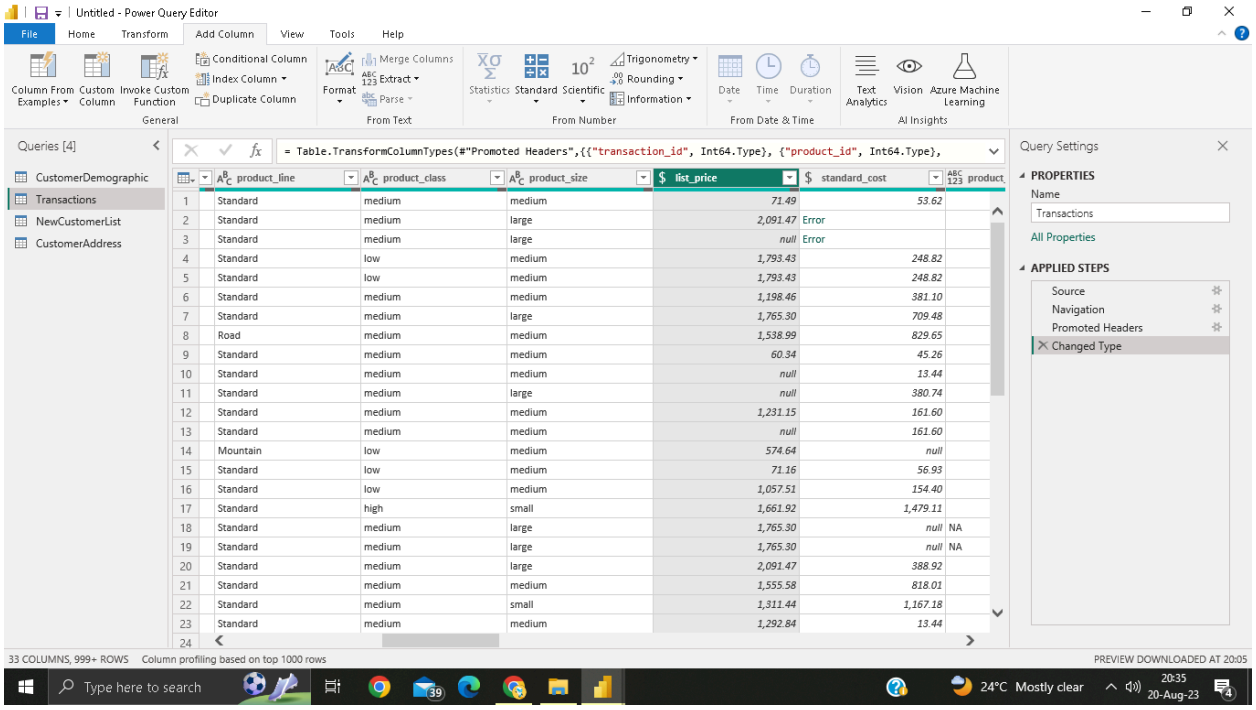


Fig.- Changed Data Type of list_price

Column Quality

Shows the percentage of rows within a column that are valid, empty, or errors. From there we can hover over the summary box (or right-click on the column quality section) for additional prompts to correct any issues.

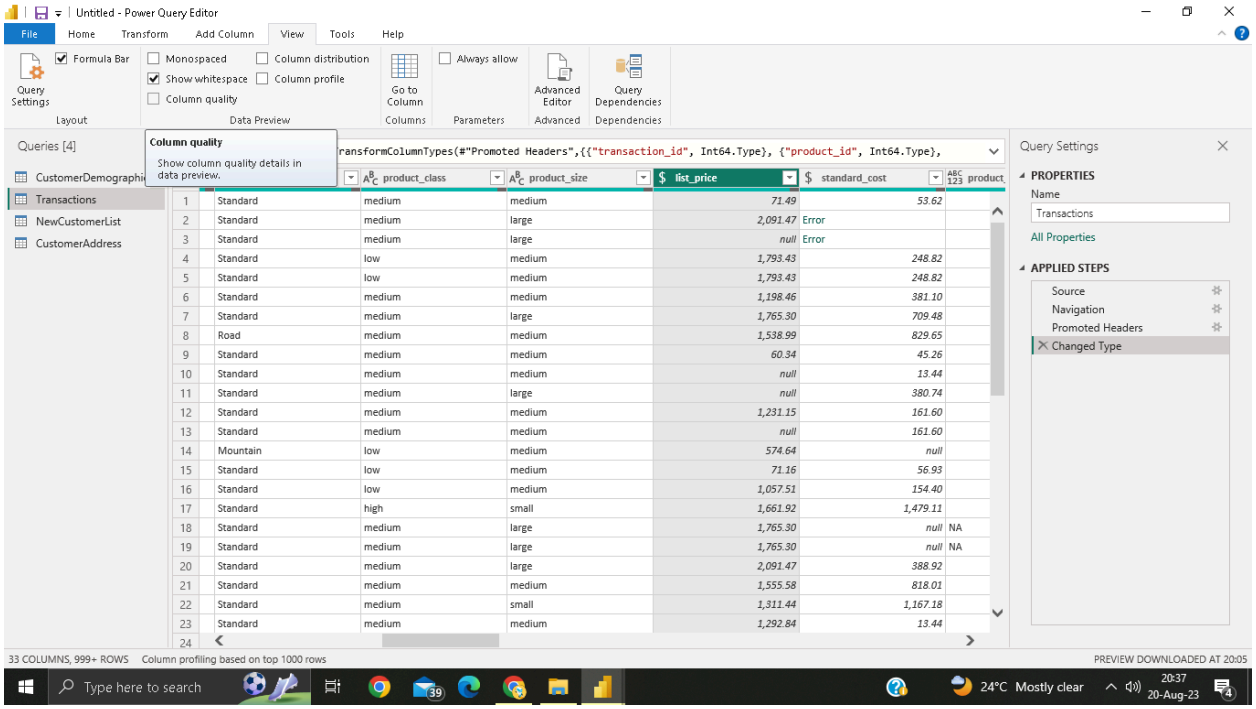


Fig.-Show column quality details in data preview

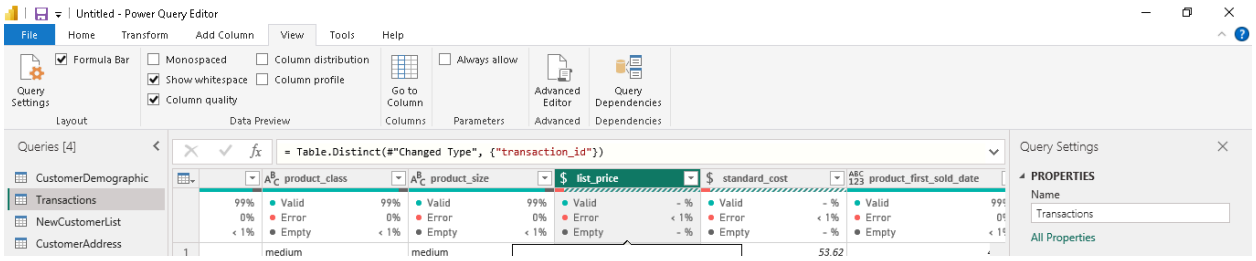


Fig.-Column quality details

Remove Duplicate

Use for removing duplicate rows and columns in Power BI. To do this, select the rows or columns that contain duplicates, then right-click and select the "Remove Duplicate Rows" or "Remove Duplicate Columns" option. This will create a new table with the duplicate rows or columns removed.

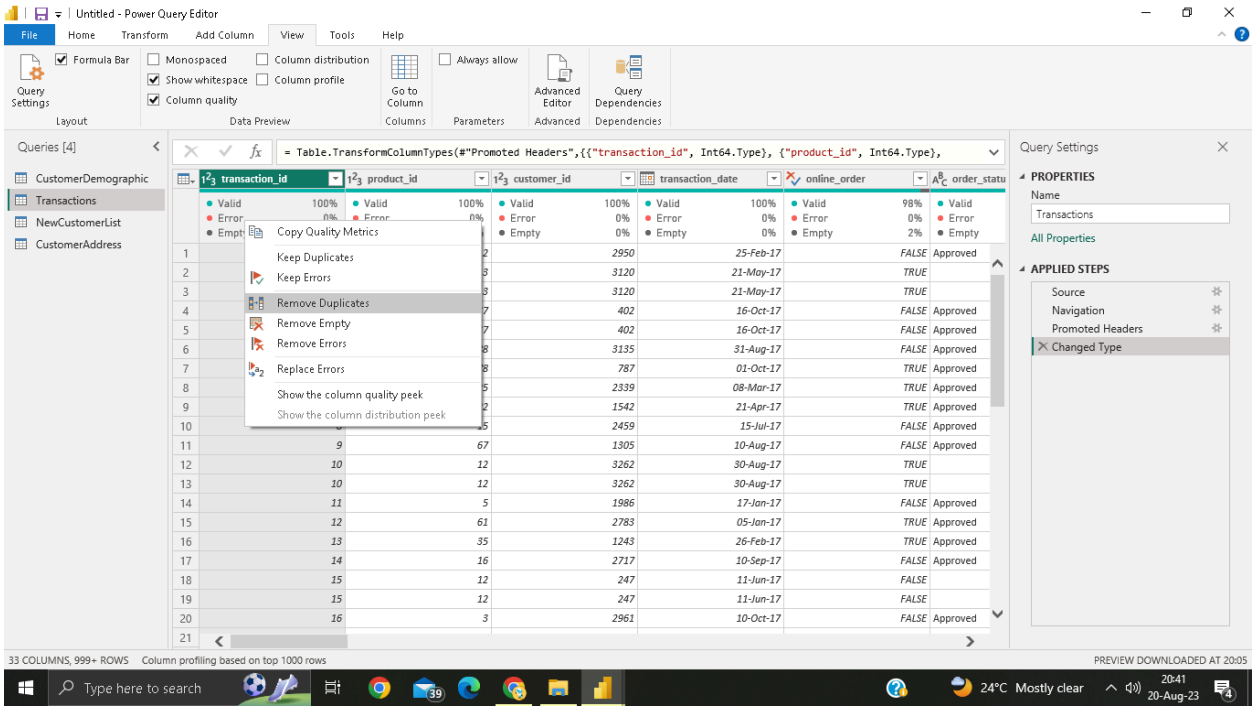


Fig.-Remove duplicates in transaction_id

Remove Errors

To remove rows with errors in Power Query, first select the column that contains errors.

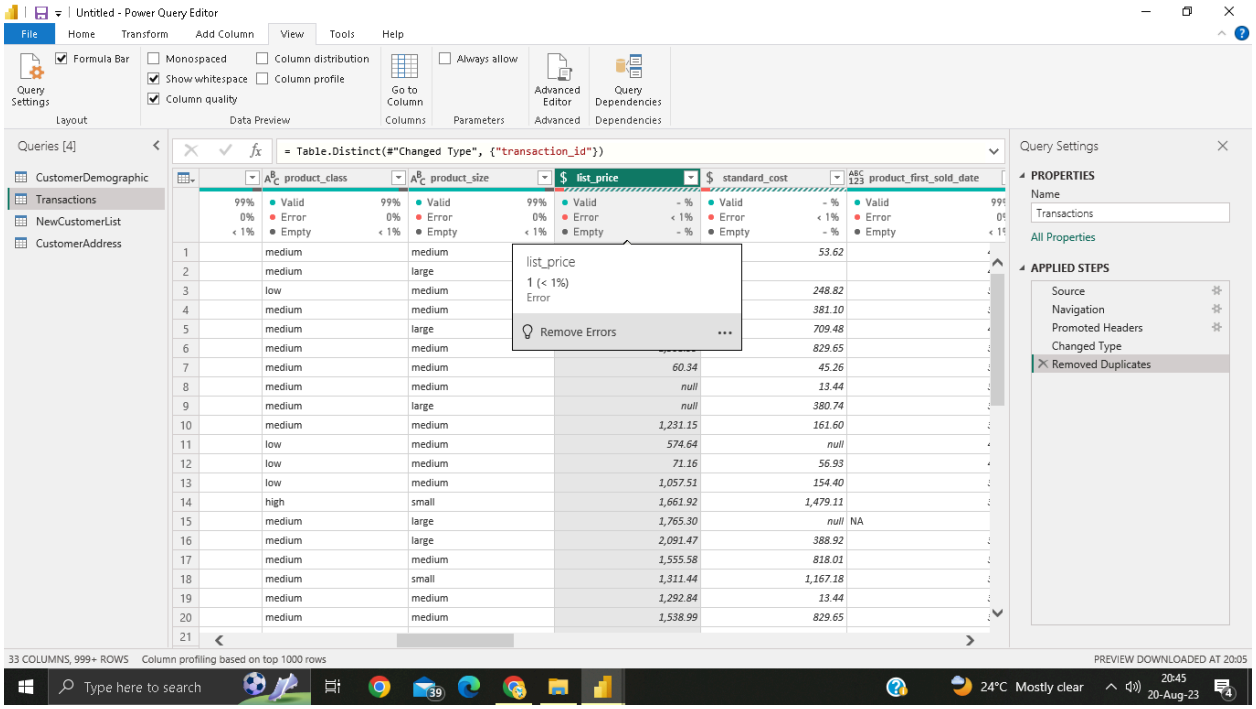


Fig.-list_price and standard_price column with Errors

From the drop-down menu, select Remove errors.

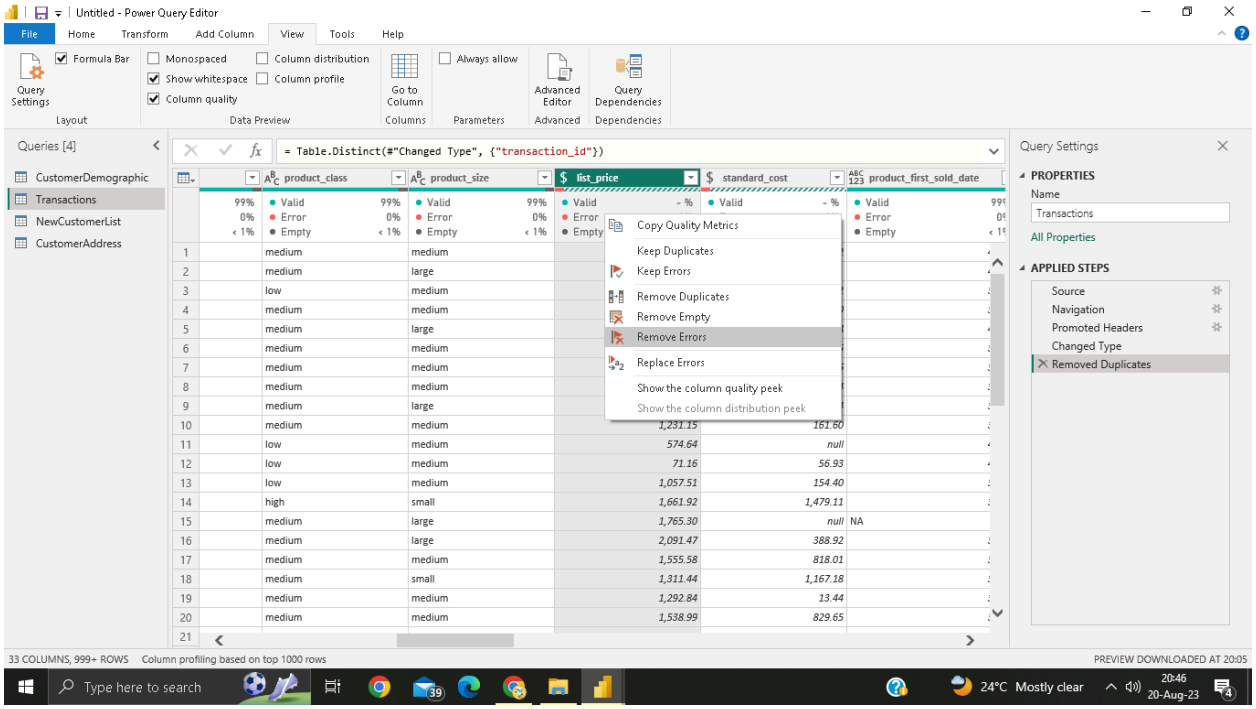


Fig.-Remove errors

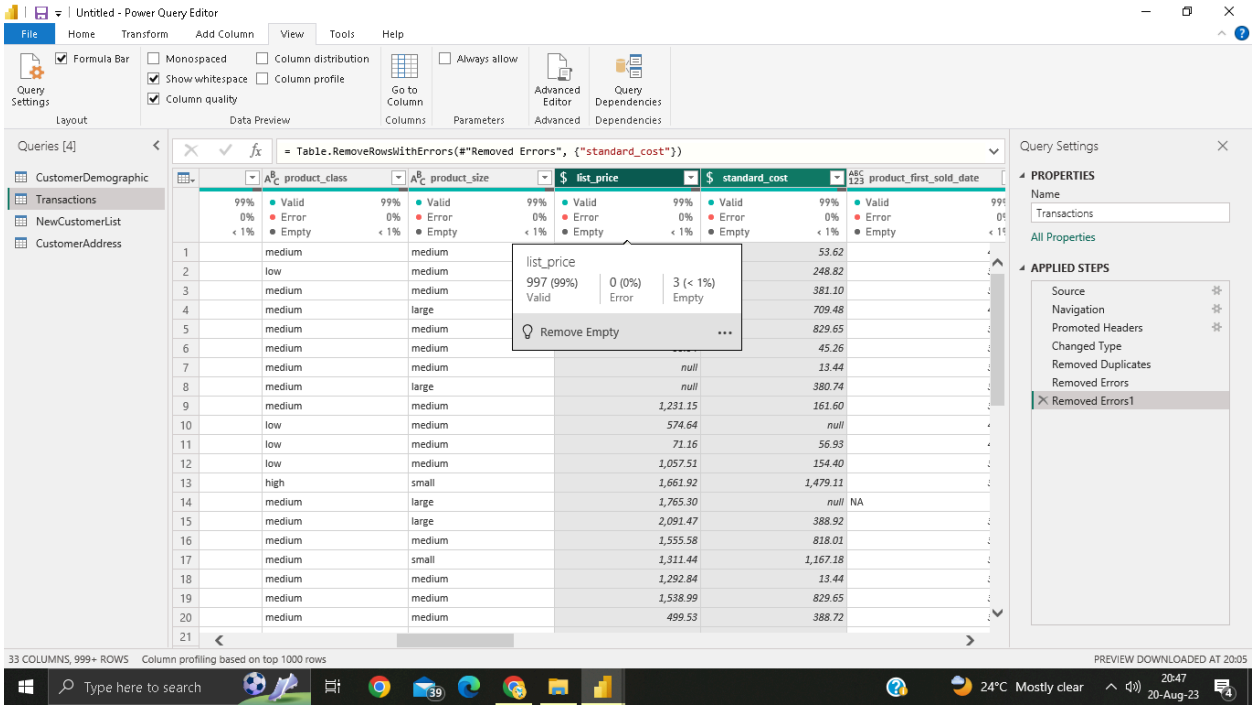


Fig.-list_price and standard_price column without Errors

Remove Empty

When you have blank values in your data table, you can use Power Query transformations to remove the row with blank values.

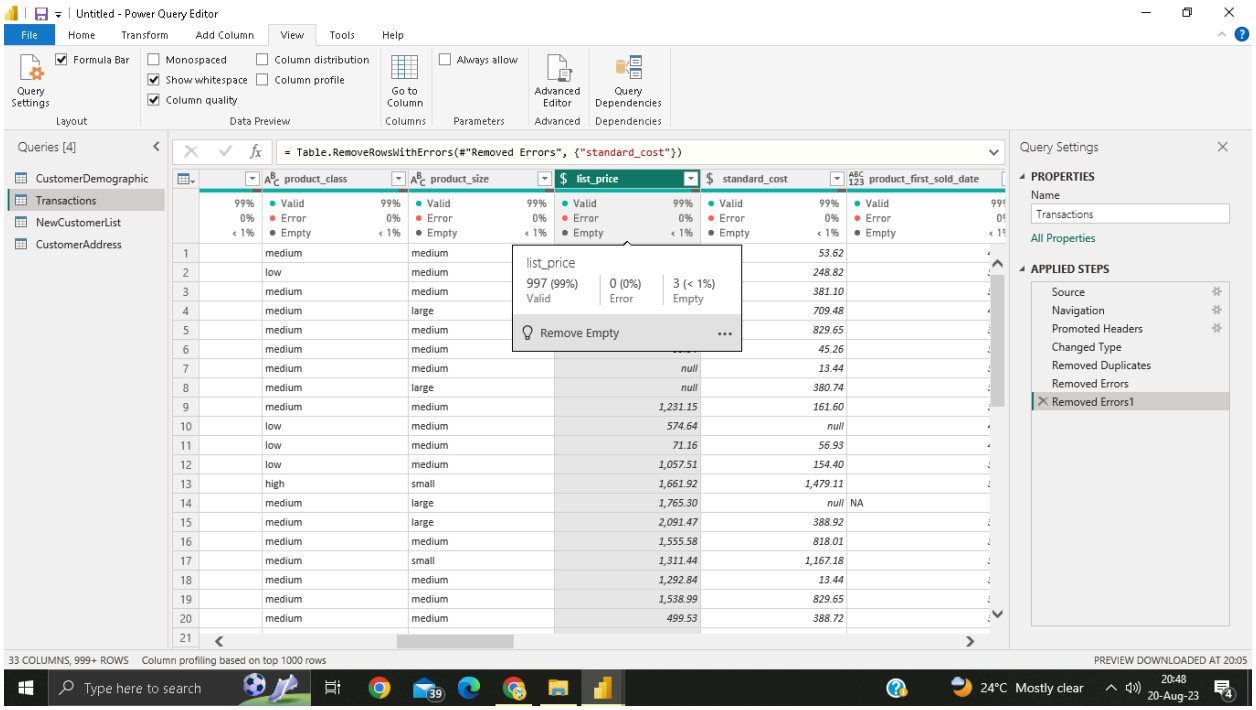


Fig.-list_price column 3 rows are empty

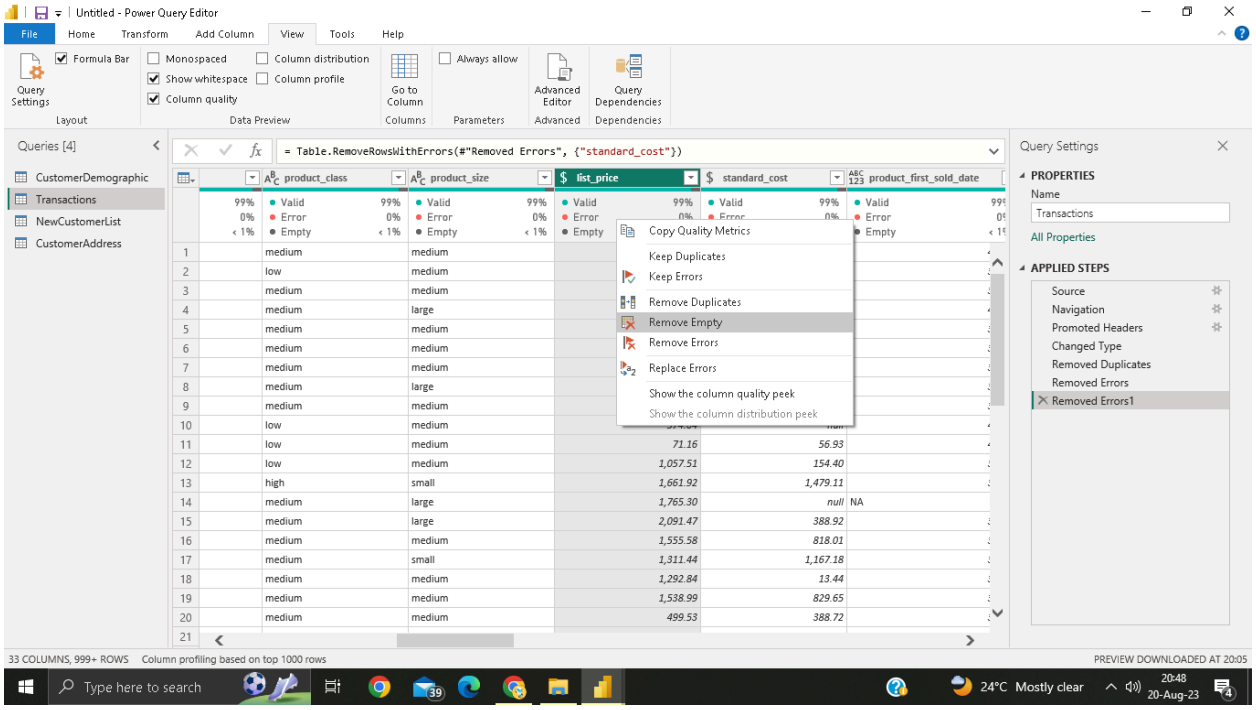


Fig.-list_price column remove empty

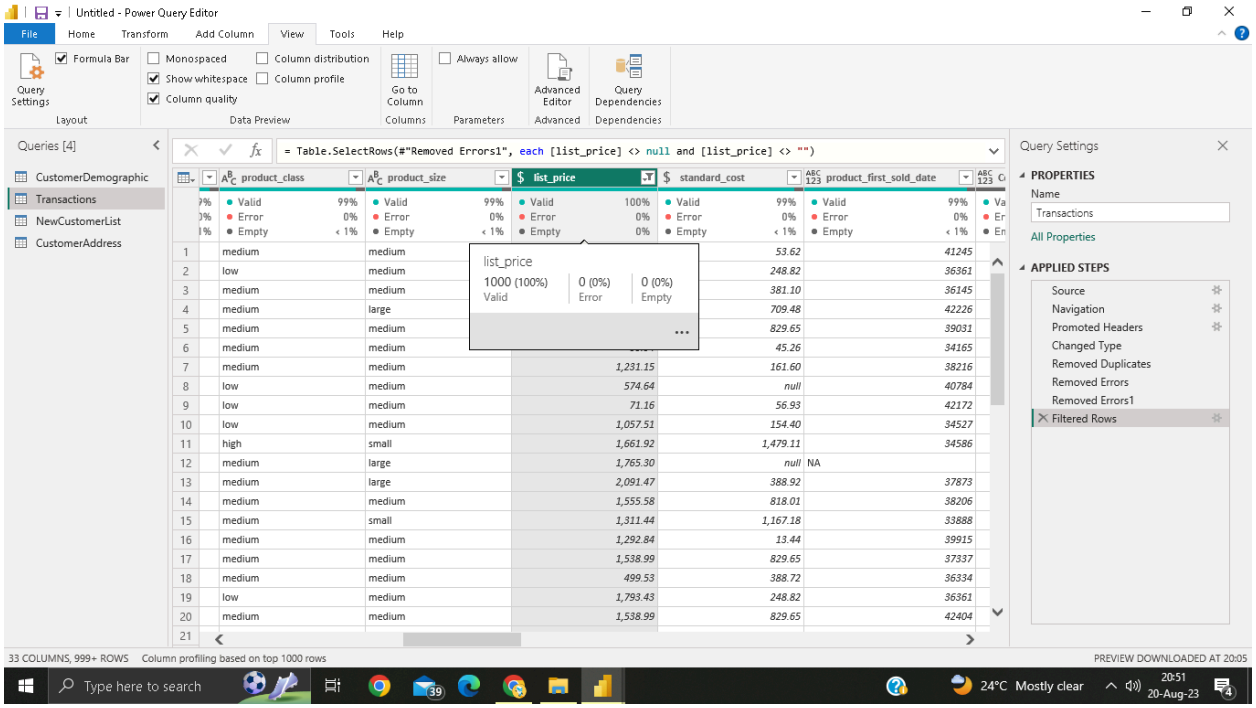


Fig.-list_price column with no empty row

Replace values

In Power Query, you can replace one value with another value in a selected column. You can replace specific values or the whole value in a cell.

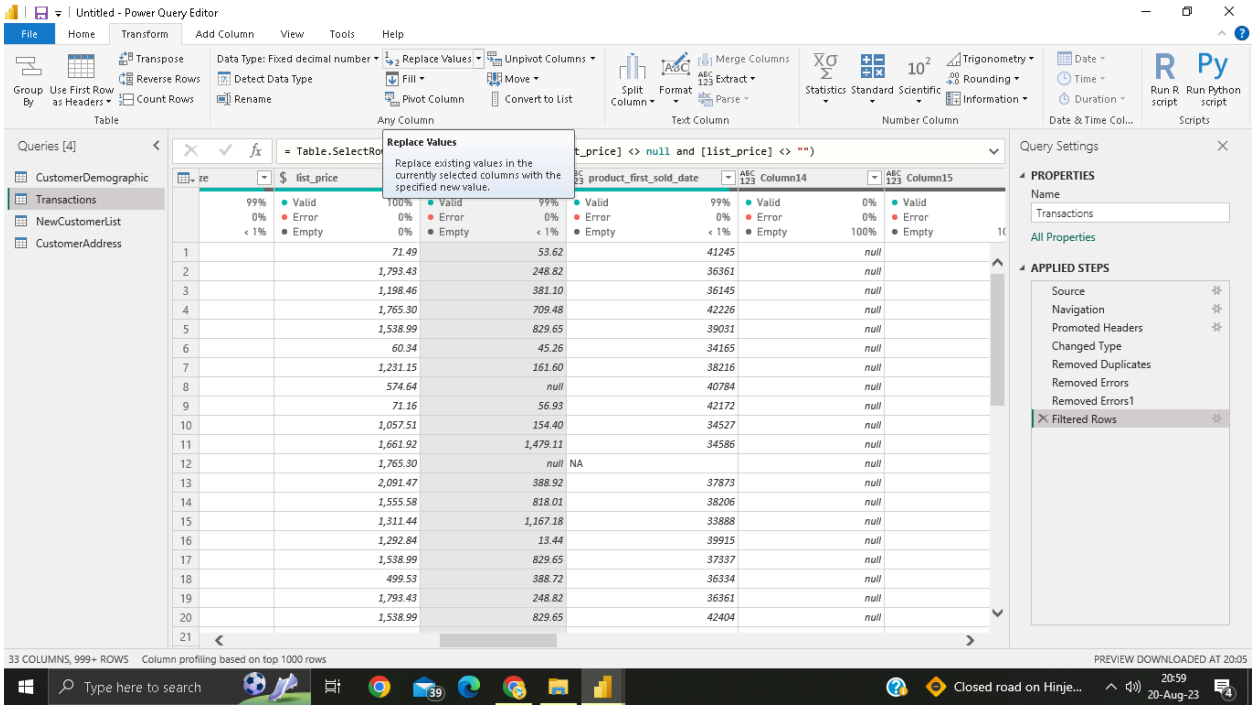


Fig.-Replace value in standard_cost column

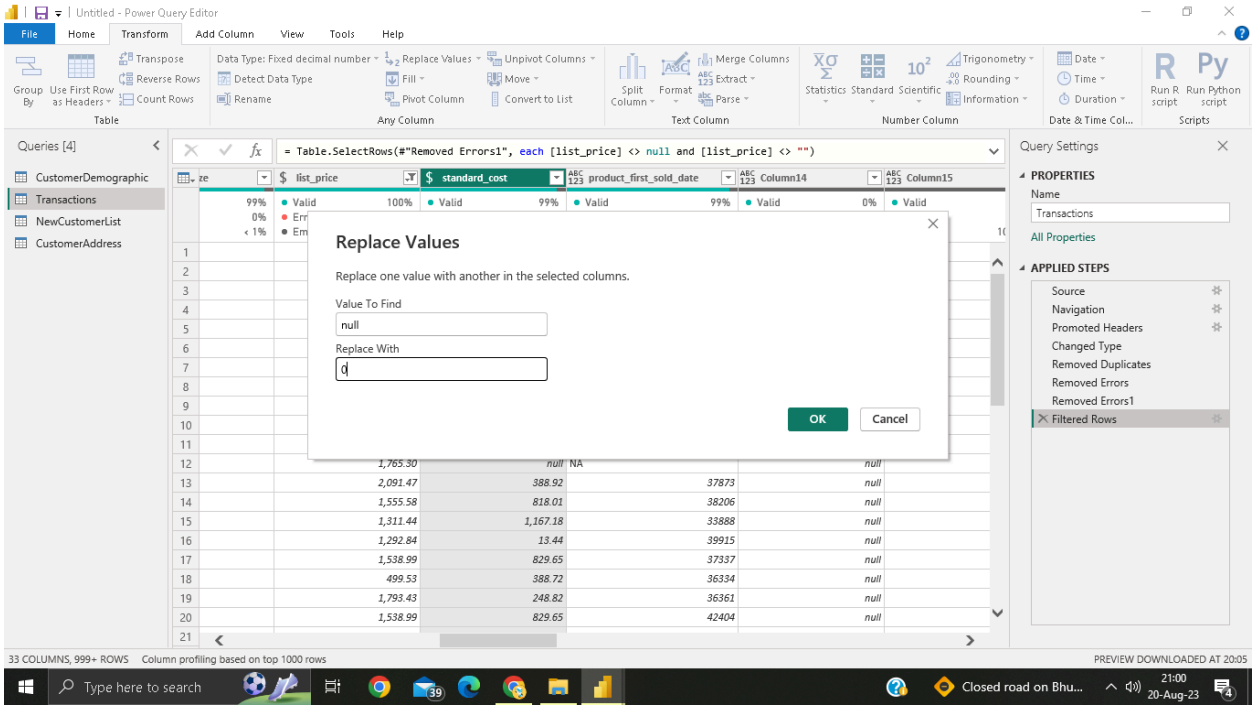


Fig.-Replace value Null with 0 in standard_cost column

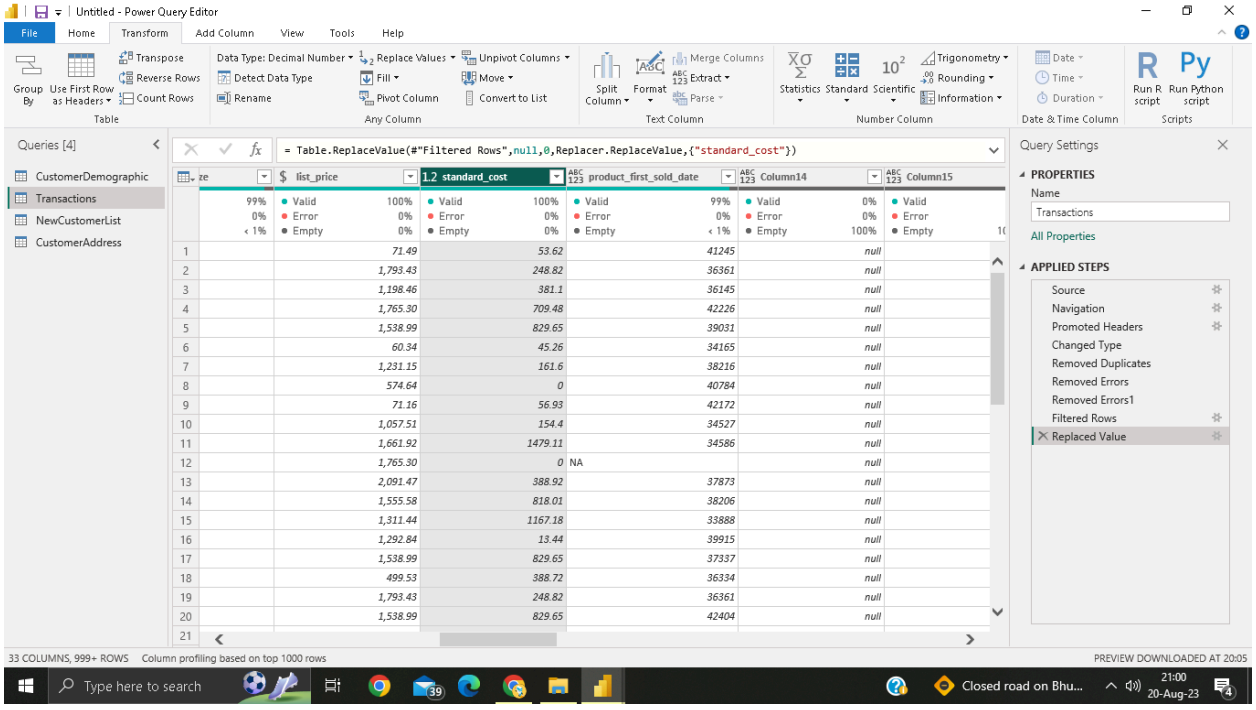


Fig.- Null value Replaced with 0 in standard_cost column

Extract Component

Extract Month & Year from a Date, In the 'Date' dialog box, select the 'Month' options. This formula creates a new text column that displays the month

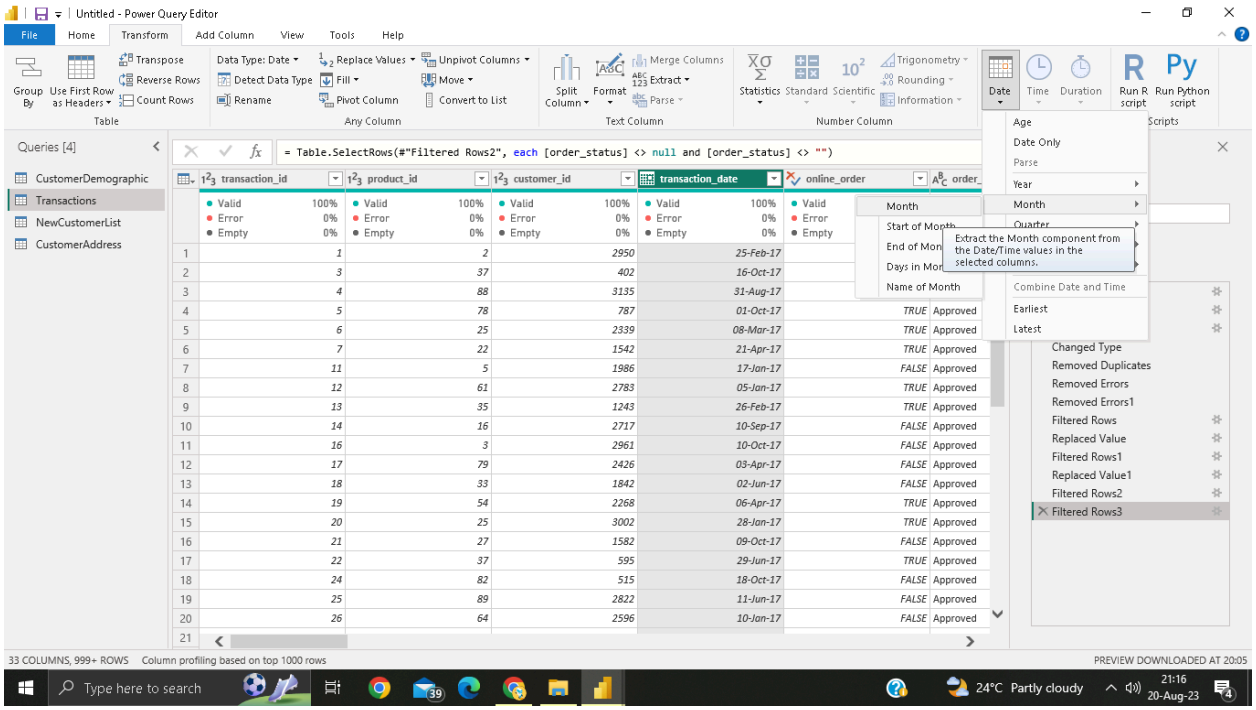


Fig.-Extract Month from transaction_date column

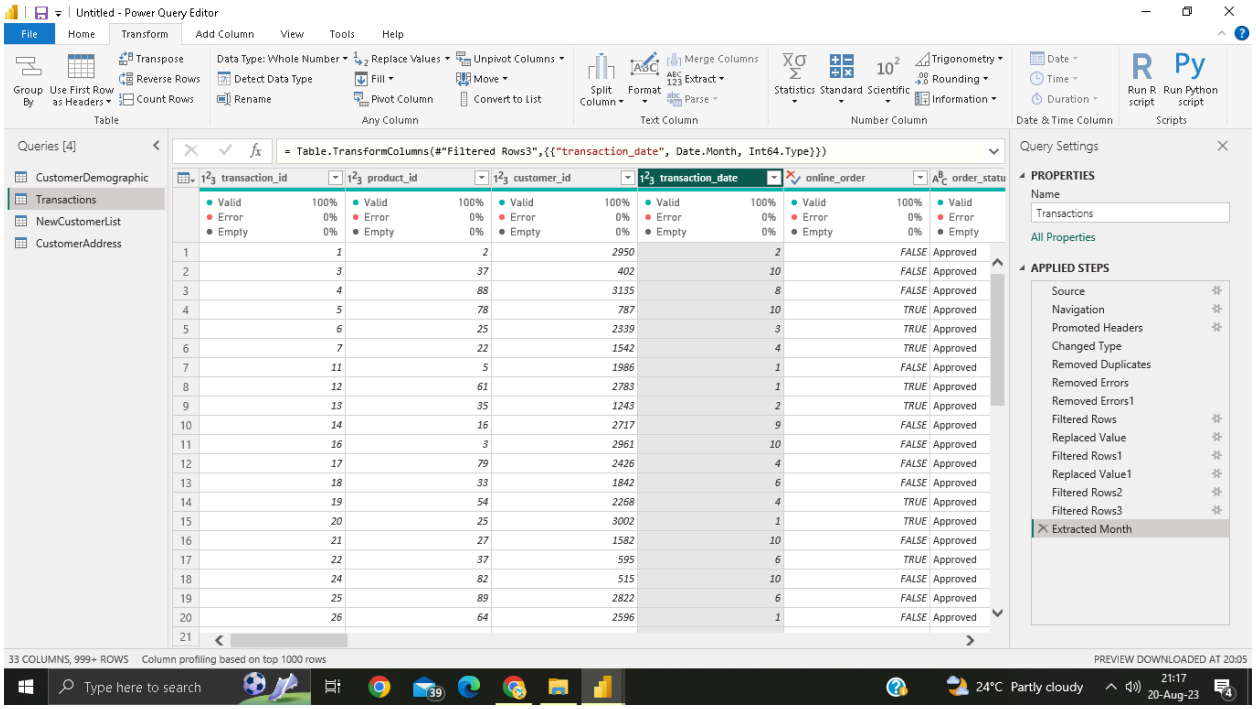


Fig.-Extracted Month from transaction_date column

Extract Month & Year from a Date, In the 'Date' dialog box, select the 'Year' options. This formula creates a new text column that displays the Year

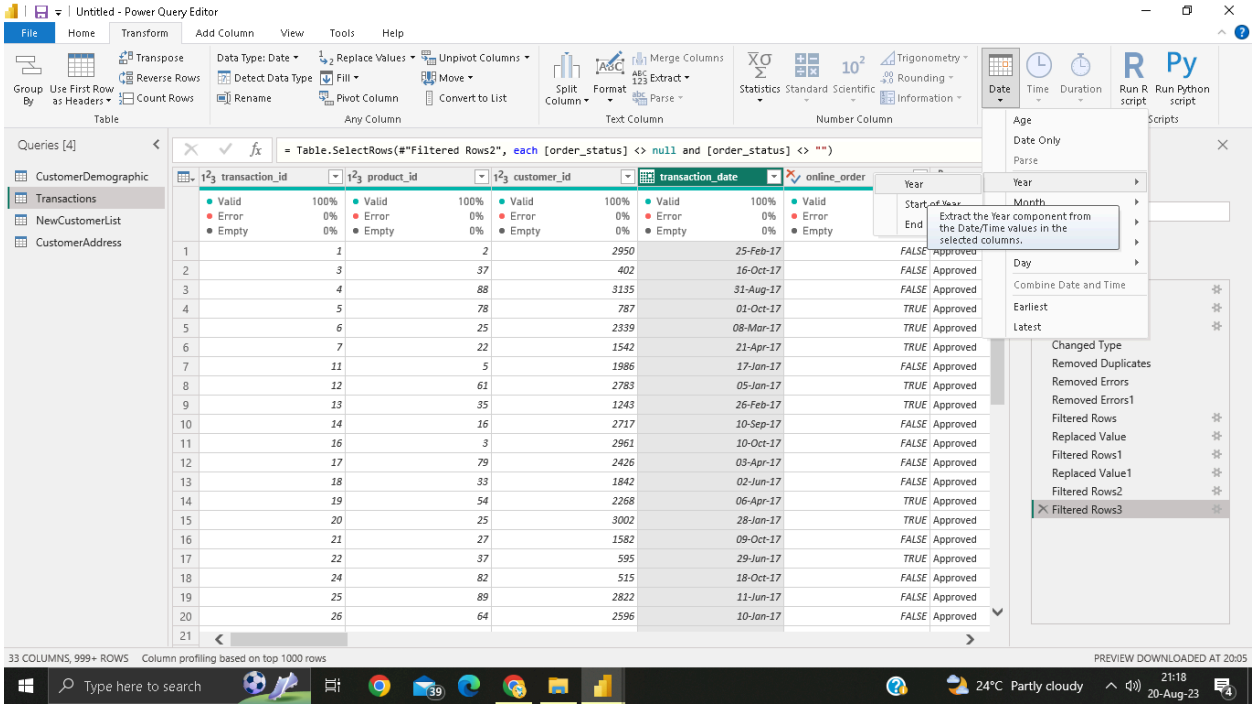


Fig.-Extract Year from transaction_date column

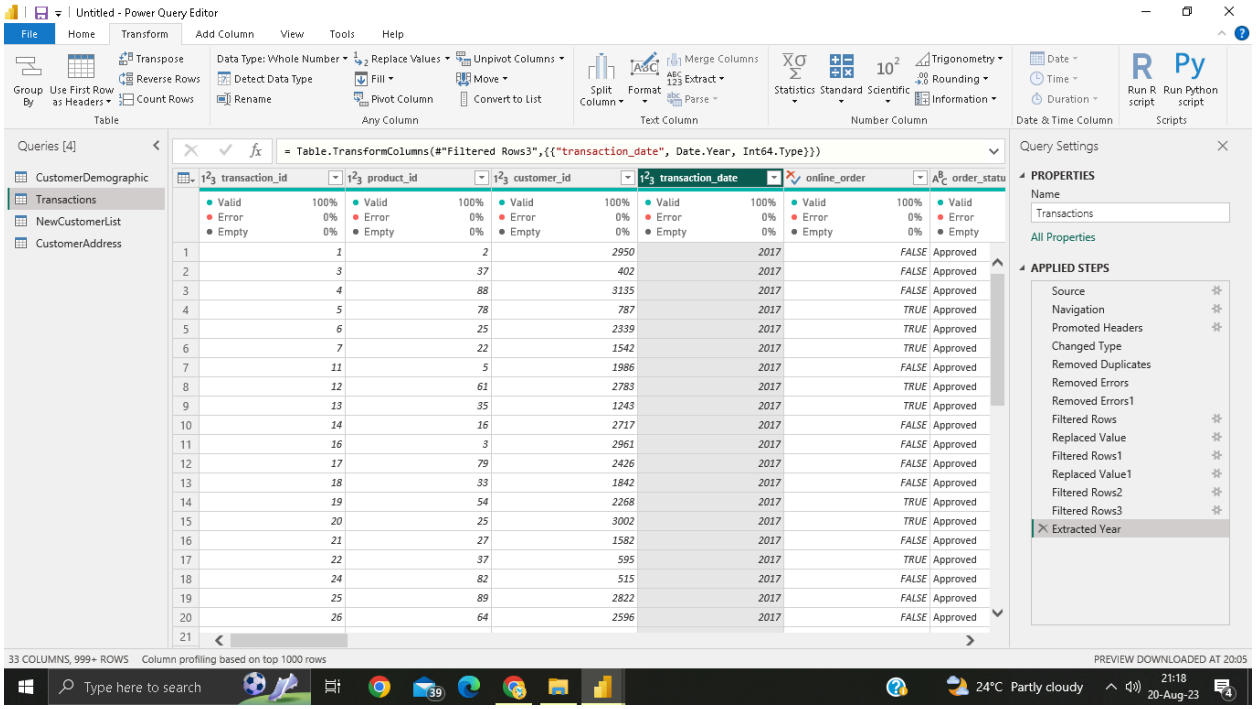


Fig.-Extracted Year from transaction_date column

Use Headers as First Row

To promote the first row to column headers, select Home > Use First Row As Headers. To demote column headers to the first row, select Home, select the arrow next to Use First Row As Headers, and then select Use Headers as First Row.

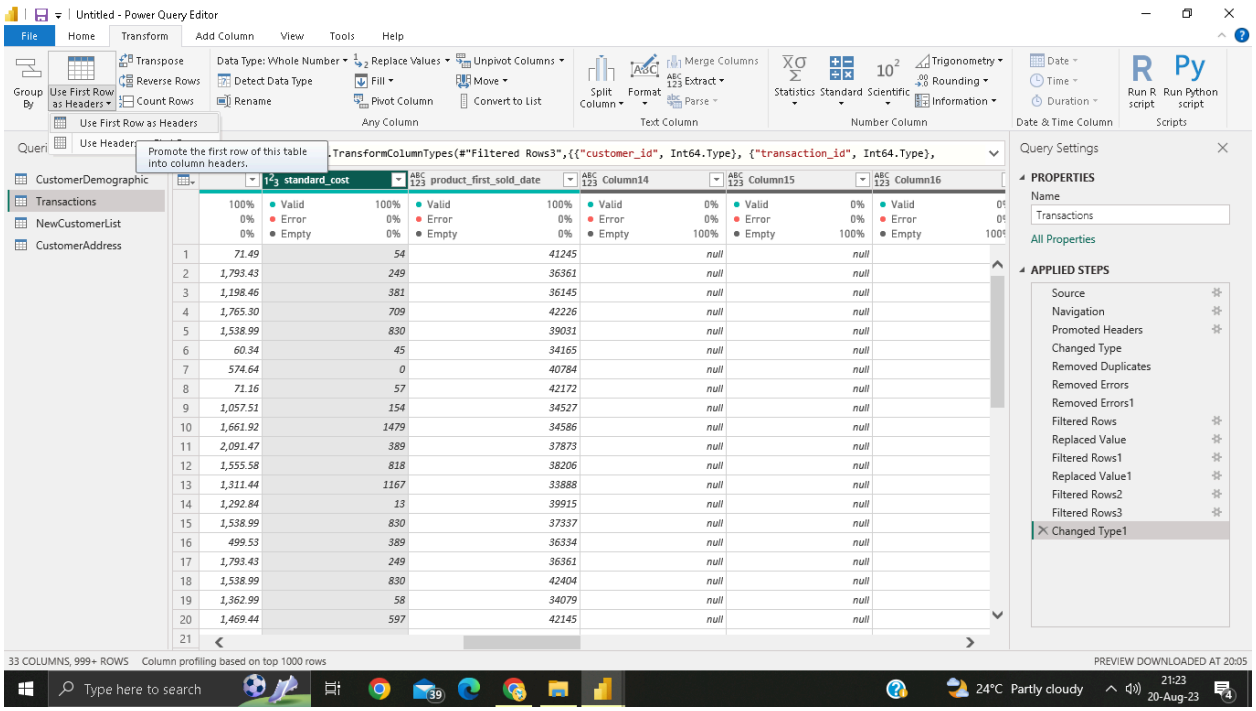


Fig.-Use First Row As Headers

Group By

The Group By feature in Power BI summarizes data based on one or more aggregate functions. You can use the Group By feature to find the average, count, min, max, or any other aggregate value for one column, based on unique values in other columns.

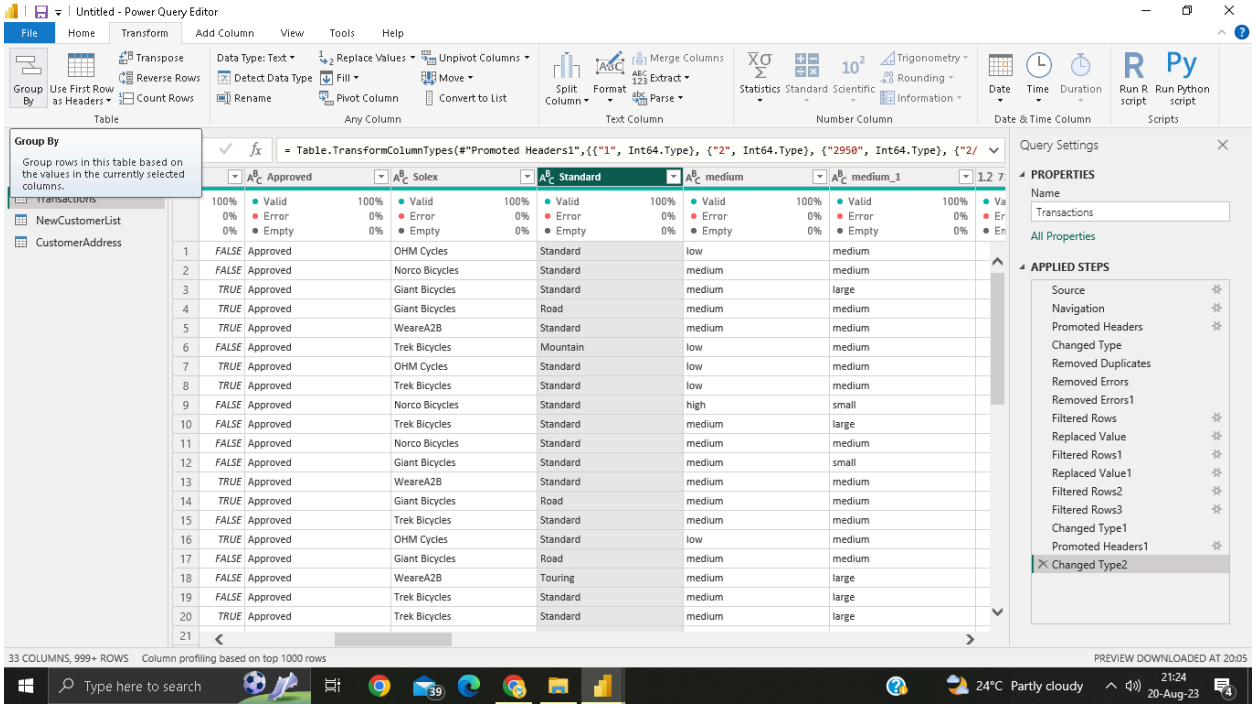


Fig.-Use group by on standard column

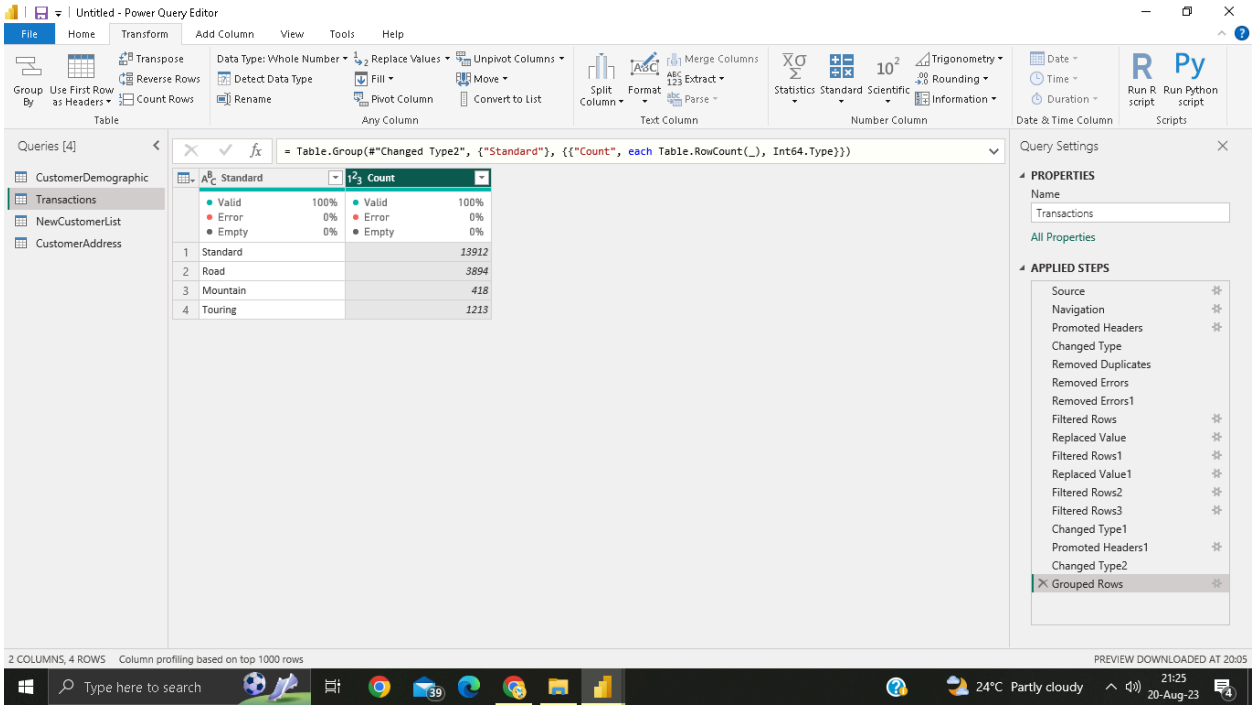


Fig.-Grouped row of standard column

Reverse rows

With Power Query, it's possible to reverse the order of rows in a table.

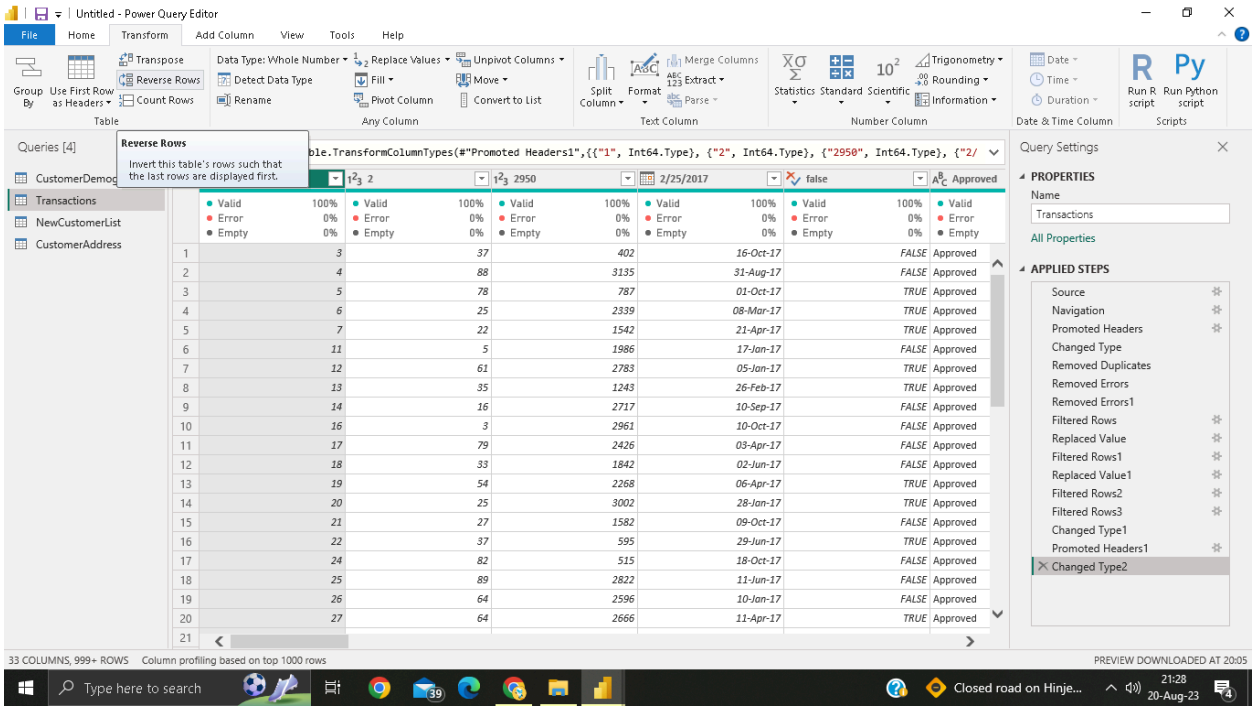


Fig.-Reverse rows of column 1

Move option

A common process when preparing data is to move columns in the dataset.

To accomplish this move, you can either select the Move option or drag and drop the column.

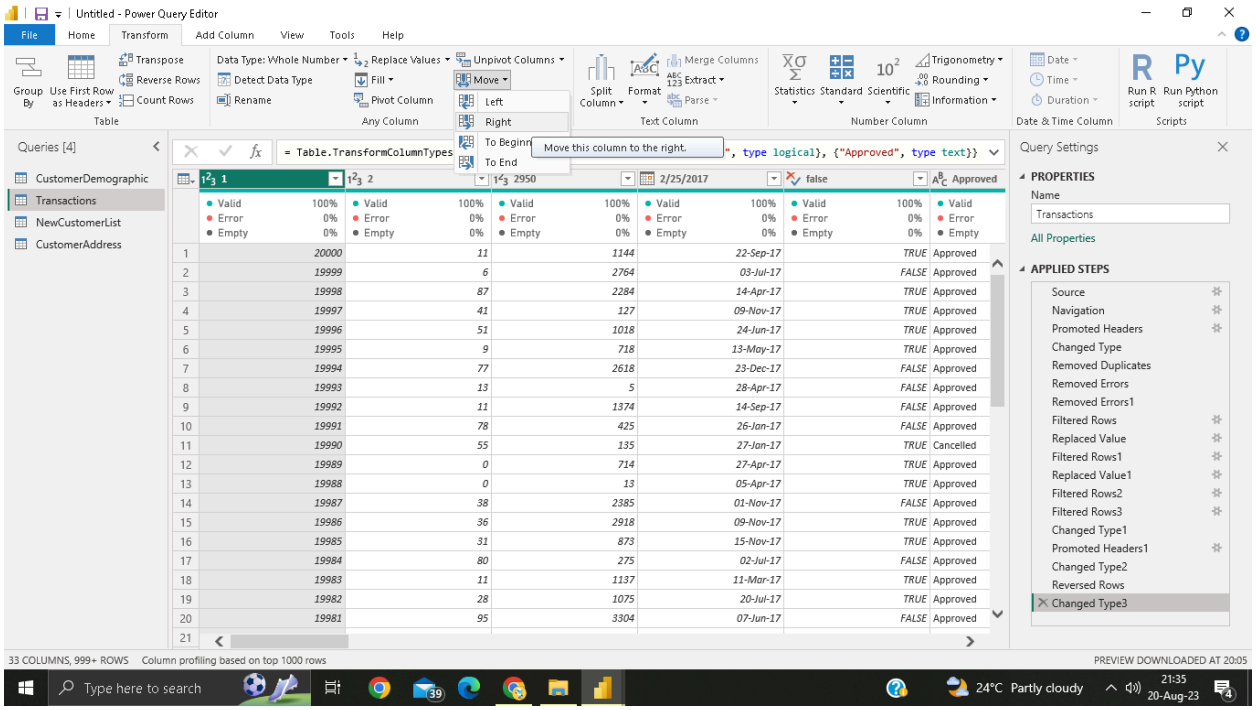


Fig.-Move selected column to right

Age

There is an Age menu option in the Power Query editor under Date. When you select a date column and use that Age option, it calculates the duration between the selected date and the current date in days.

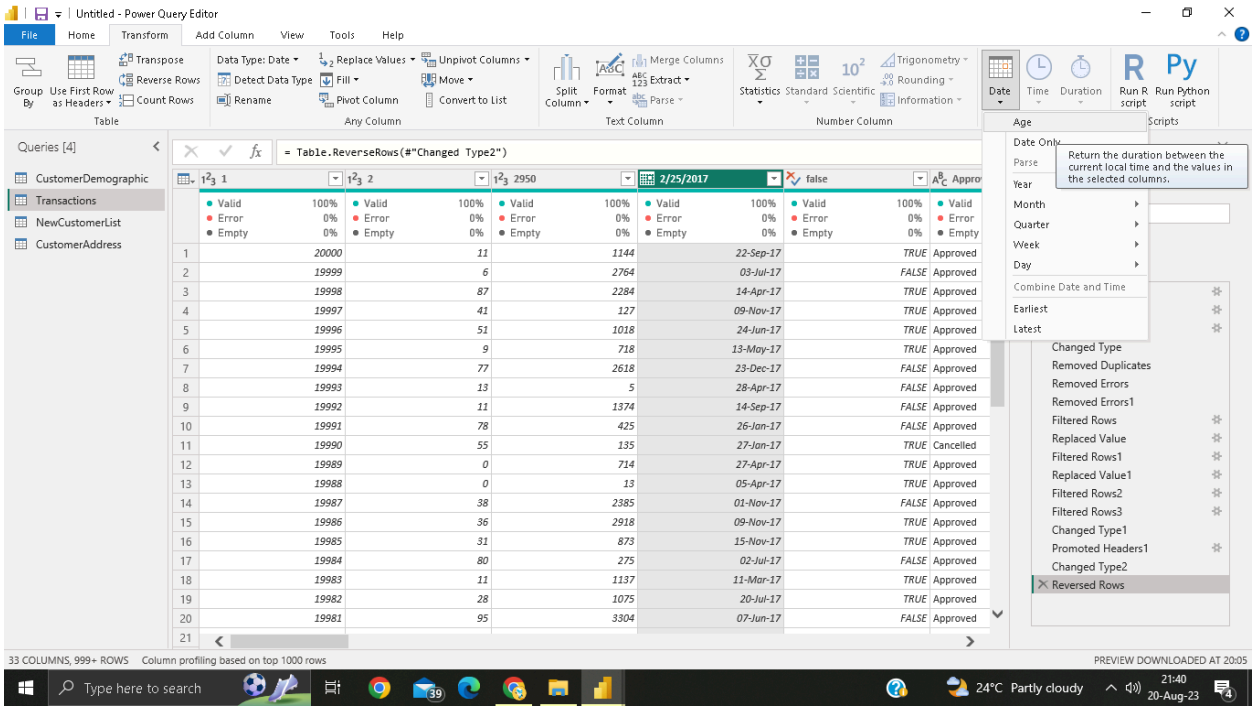


Fig.-Calculates the duration between the selected date

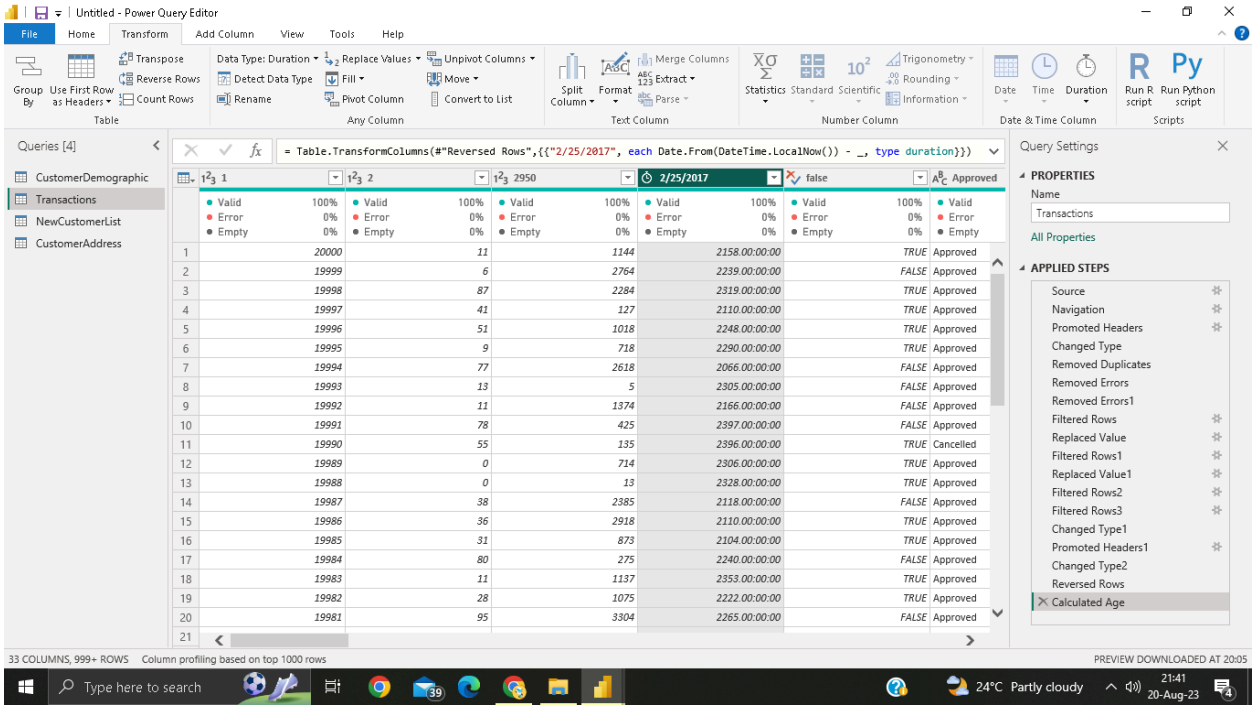


Fig.-Calculated duration between the selected date

Sort data

We can sort data to arrange in:
Ascending order (text is A-Z, number is smallest to largest, and date is oldest to newest).
Descending order (text is Z-A, number is highest to lowest, and date is newest to oldest).

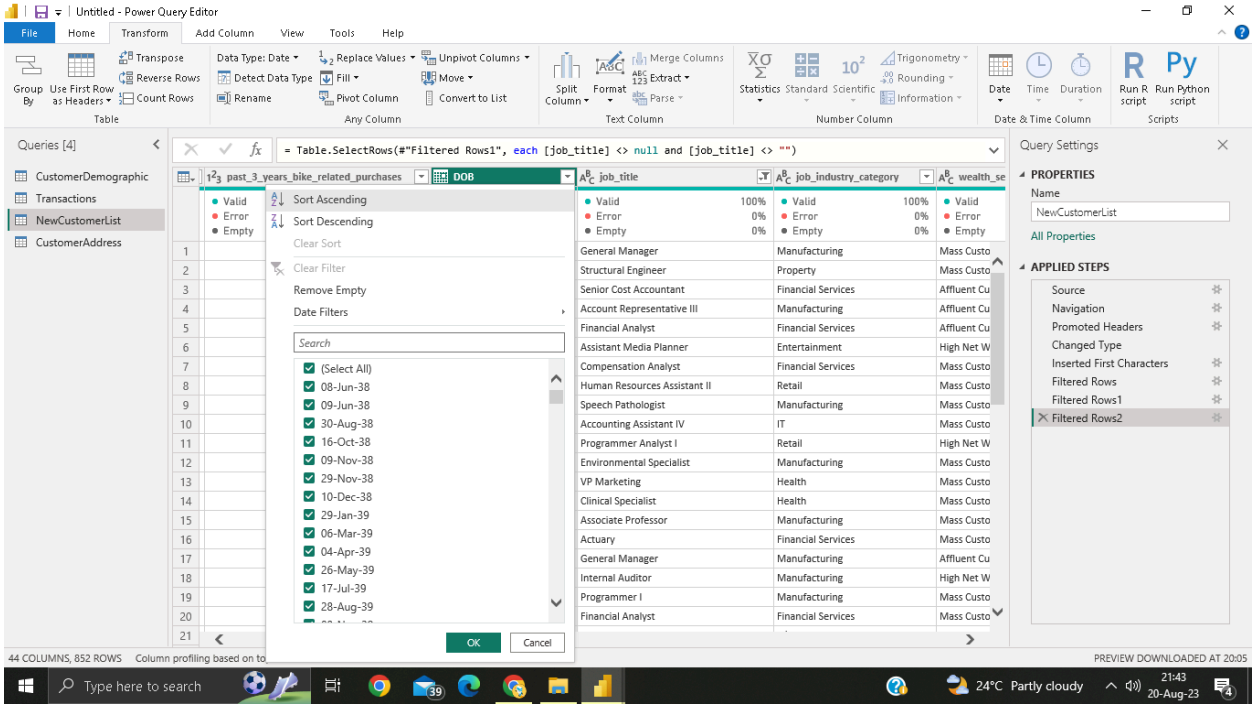


Fig.-Sort Ascending order of DOB column

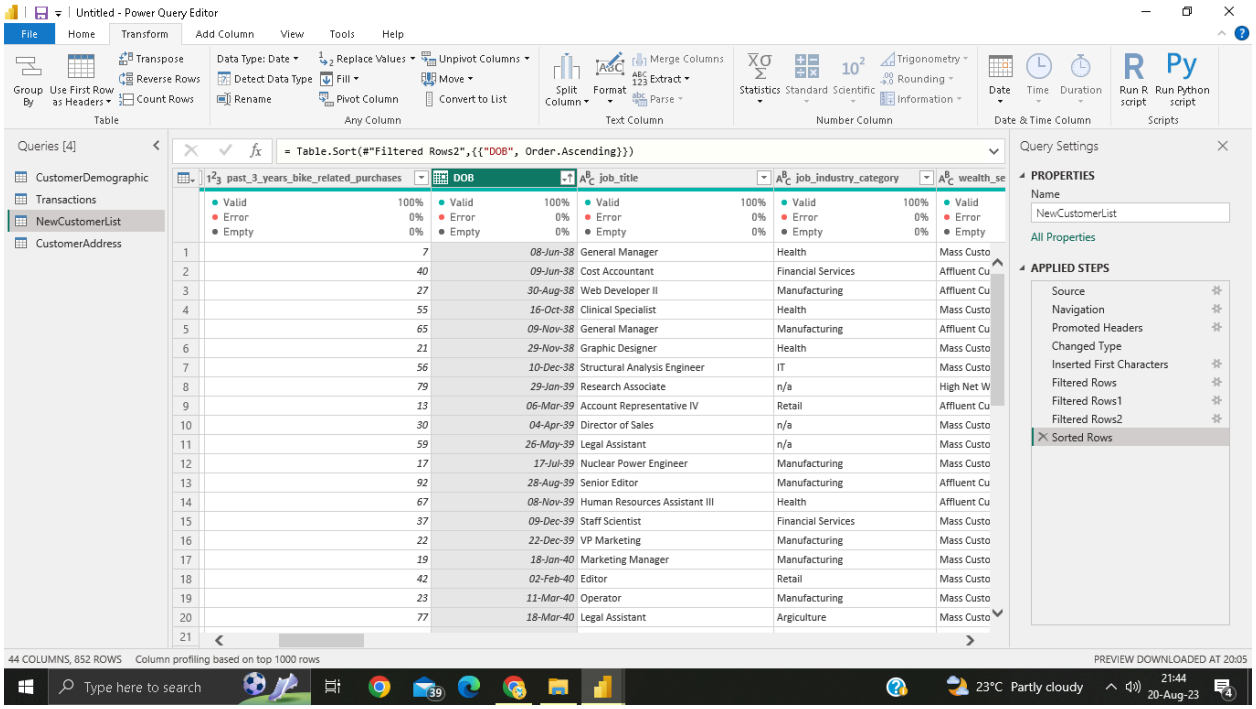


Fig.-Sort Ascending order of DOB

Save your work

When our query is where we want , select Close & Apply from Power Query Editor's File menu. This action applies the changes and closes the editor.

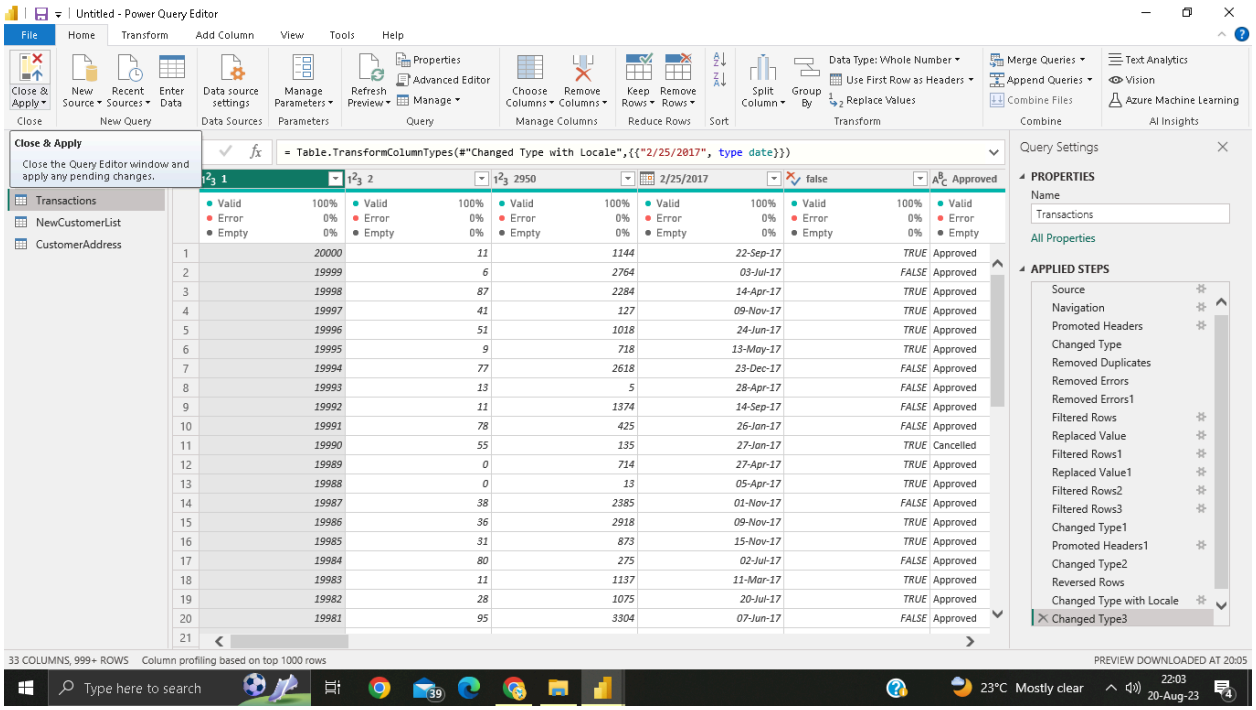


Fig.-Close the editor

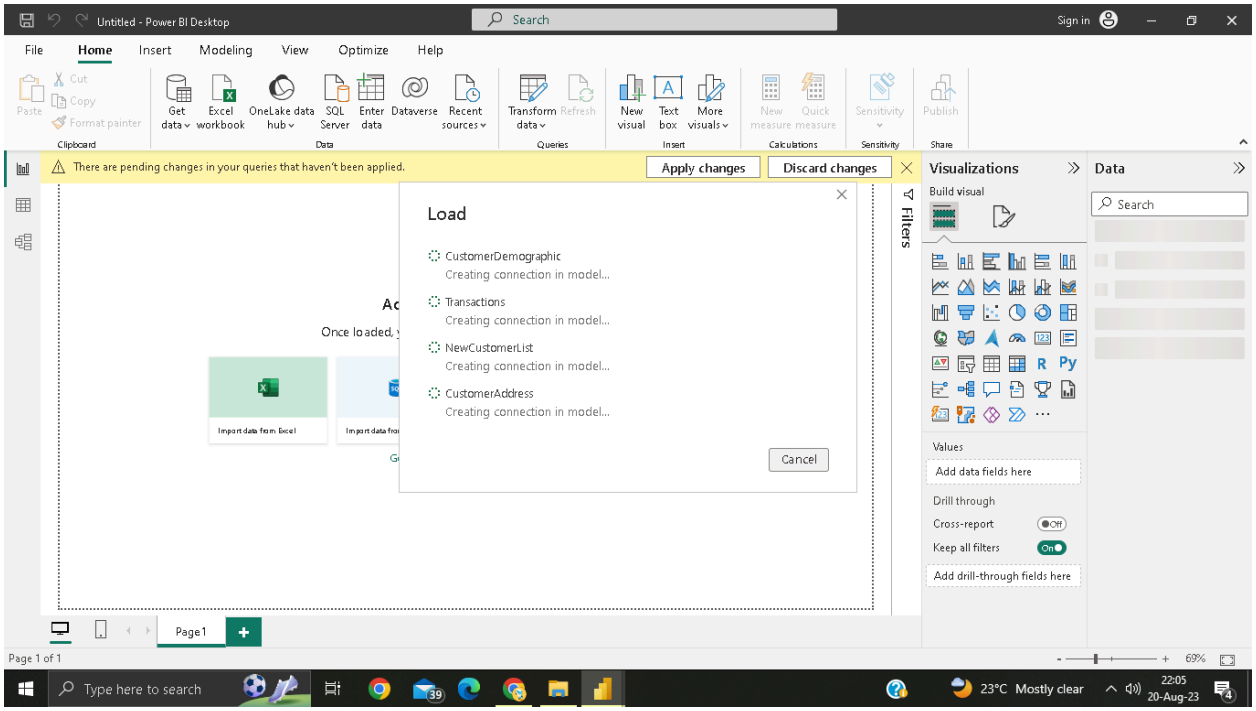


Fig.-Creating model

Model view

Get a graphical representation of the relationships that are established in our data model, and manage or modify them as needed.

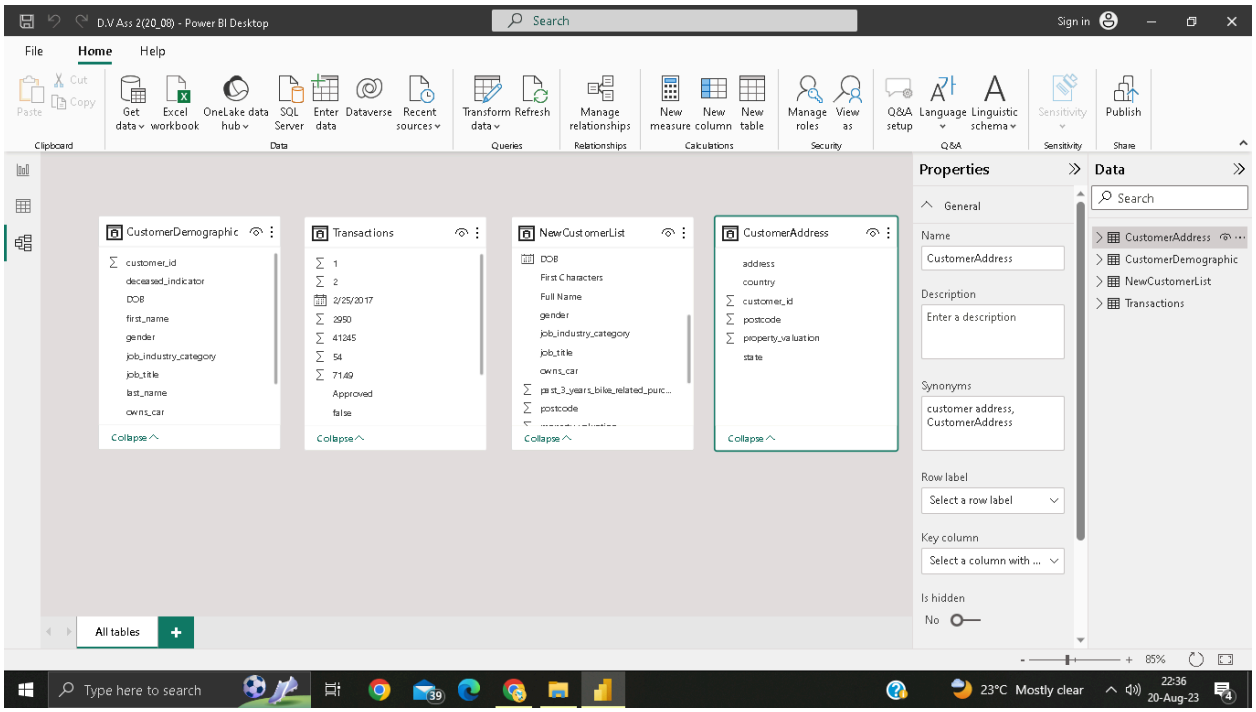


Fig.-Model view