DV Analytics
Transforming You

Power Query Deep Dive

Venkata Reddy Konasani

Contents

- Power Query
- M-language
- Useful commands in M-Language
- Data model

Mastering Power BI

- If you want to master Power BI, there are three aspects that we need to be very good at
  - Power Query M Language
  - DAX
  - Story telling.
- The storytelling part involves creativity, but M Language and DAX can be mastered quickly.
- Lets take one more case study to see in-depth concepts of M-Language and DAX

Case Study :
E-commerce Product Sales Analysis

E-commerce Product Sales Analysis

Step-1 : Understand the Problem Statement

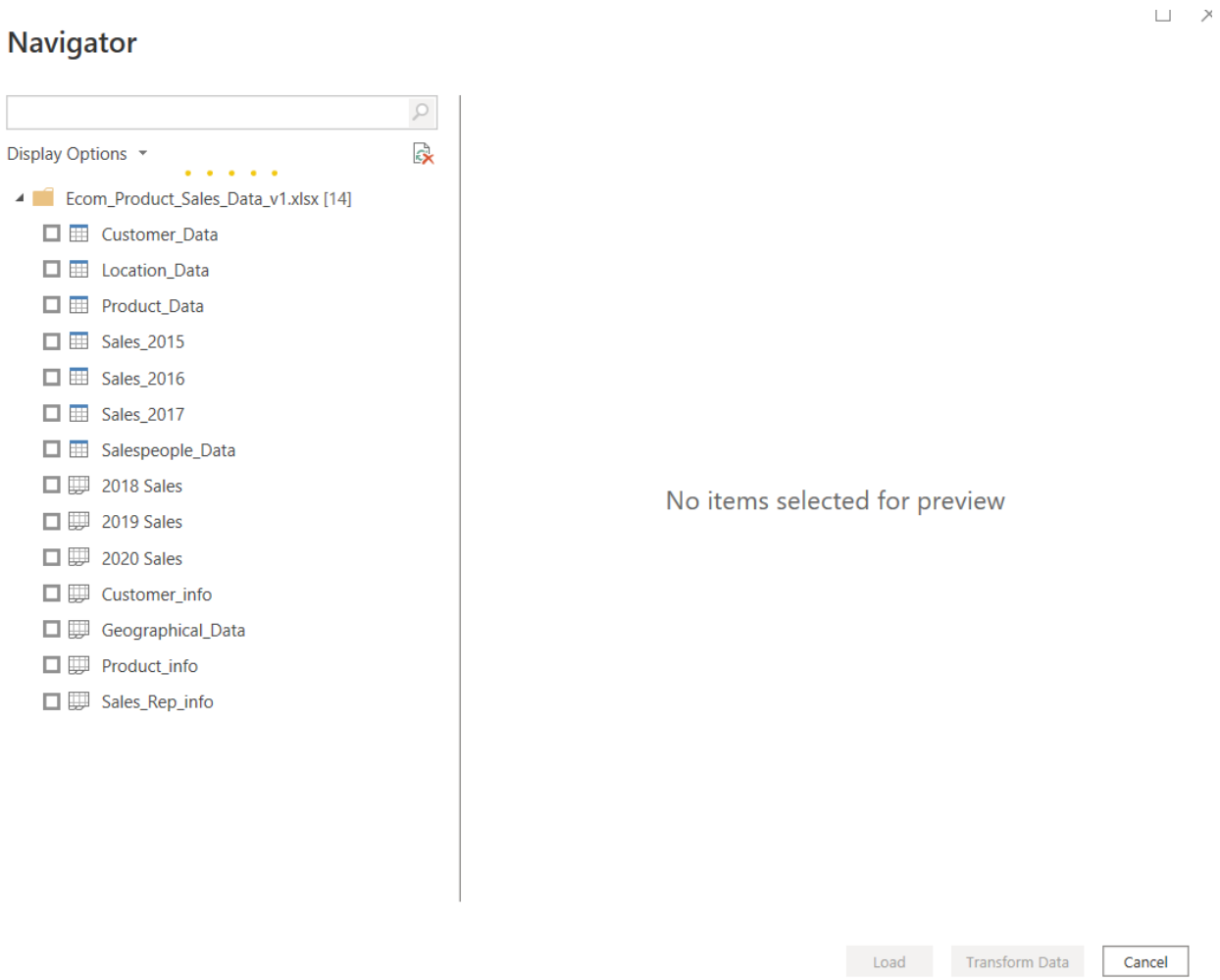
Problem Statement

- Analyze Sales and Profits in various scenarios.
- Analyse Ecommerce data and find the hidden patters that will cause high or low sales in a category
- Analyze the profits and show how profit varies in different segments

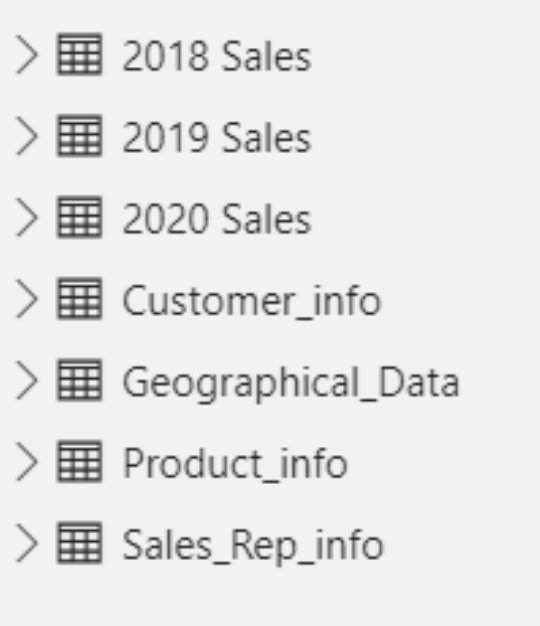
E-commerce Product Sales Analysis

Step-2 : Load the Data and get the basic details

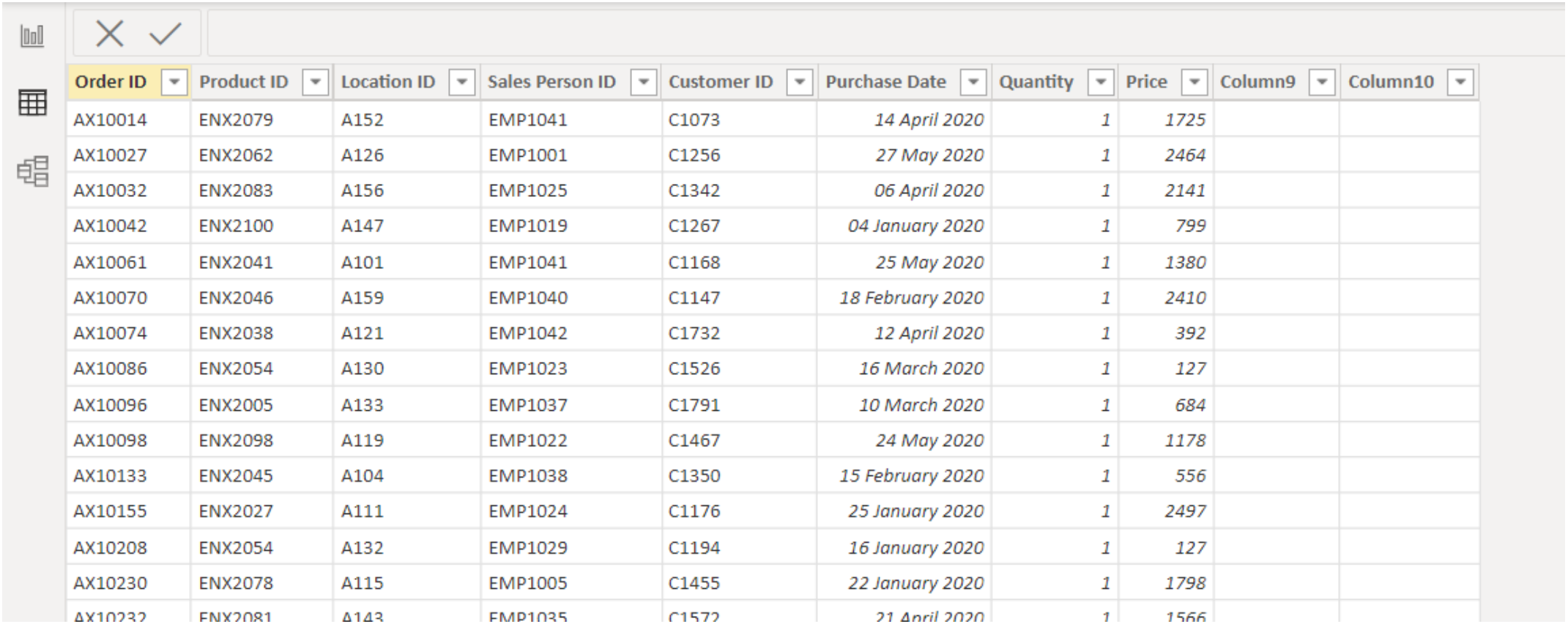
Load the Data and get the basic details



Data Set


2018 Sales
2019 Sales
2020 Sales
Customer_info
Geographical_Data
Product_info
Sales_Rep_info

Sales Table



Order ID	Product ID	Location ID	Sales Person ID	Customer ID	Purchase Date	Quantity	Price	Column9	Column10
AX10014	ENX2079	A152	EMP1041	C1073	14 April 2020	1	1725		
AX10027	ENX2062	A126	EMP1001	C1256	27 May 2020	1	2464		
AX10032	ENX2083	A156	EMP1025	C1342	06 April 2020	1	2141		
AX10042	ENX2100	A147	EMP1019	C1267	04 January 2020	1	799		
AX10061	ENX2041	A101	EMP1041	C1168	25 May 2020	1	1380		
AX10070	ENX2046	A159	EMP1040	C1147	18 February 2020	1	2410		
AX10074	ENX2038	A121	EMP1042	C1732	12 April 2020	1	392		
AX10086	ENX2054	A130	EMP1023	C1526	16 March 2020	1	127		
AX10096	ENX2005	A133	EMP1037	C1791	10 March 2020	1	684		
AX10098	ENX2098	A119	EMP1022	C1467	24 May 2020	1	1178		
AX10133	ENX2045	A104	EMP1038	C1350	15 February 2020	1	556		
AX10155	ENX2027	A111	EMP1024	C1176	25 January 2020	1	2497		
AX10208	ENX2054	A132	EMP1029	C1194	16 January 2020	1	127		
AX10230	ENX2078	A115	EMP1005	C1455	22 January 2020	1	1798		
AX10232	ENX2081	A143	EMP1035	C1572	21 April 2020	1	1566		

Geographical table

Location ID	Name	County	State Code	State	Type	Latitude	Longitude	Area Code	Population	Households	Median Income	Land Area	Water Area	Time Zone
A100	Anaheim	Orange County	CA	California	City	33.83529	-117.9145	657	350742	99670	60752	129438076	2457142	America/Lo
A101	Antioch	Contra Costa County	CA	California	City	38.00492	-121.80579	925	110542	33718	64329	76173197	1842711	America/Lo
A102	Bakersfield	Kern County	CA	California	City	35.37329	-119.01871	661	373640	112439	57095	385366784	3809676	America/Lo
A103	Berkeley	Alameda County	CA	California	City	37.87159	-122.27275	510	120972	45917	66237	27106077	18715614	America/Lo
A104	Burbank	Los Angeles County	CA	California	City	34.18084	-118.30897	818	105319	41361	66076	44947219	94286	America/Lo
A105	Carlsbad	San Diego County	CA	California	City	33.15809	-117.35059	760	113453	42791	90597	97713477	3595655	America/Lo
A106	Chula Vista	San Diego County	CA	California	City	32.64	-117.08417	619	265757	78066	65185	128544440	6380135	America/Lo
A107	Clovis	Fresno County	CA	California	City	36.82523	-119.70292	559	104180	34512	62666	62693928	0	America/Lo
A108	Concord	Contra Costa County	CA	California	City	37.97798	-122.03107	925	128667	45409	68318	79108534	0	America/Lo
A109	Corona	Riverside County	CA	California	City	33.87529	-117.56644	951	164226	48156	74149	102233537	211897	America/Lo
A110	Costa Mesa	Orange County	CA	California	City	33.64113	-117.91867	949	113204	40908	66459	40723584	211253	America/Lo
A111	Daly City	San Mateo County	CA	California	City	37.70583	-122.46194	650	106562	31137	74449	19788422	0	America/Lo
A112	Downey	Los Angeles County	CA	California	City	33.94001	-118.13257	562	114219	32738	62897	32136795	414376	America/Lo

Fields

Search

- 2018 Sales
- 2019 Sales
- 2020 Sales
- Customer_info
- Geographical_Data
- Product_info
- Sales_Rep_info

Customer Information

Customer ID	Customer Name	Age
C1000	Jesse Evans	34
C1001	Victor Ramos	35
C1002	Mark Montgomery	45
C1003	Dennis Morris	31
C1004	Gregory Simmons	35
C1005	Jeremy Vasquez	30
C1006	Anthony Simpson	39
C1007	Ernest Rivera	32
C1008	Victor Martinez	47
C1009	Bobby Burton	43
C1010	Bruce Porter	43
C1011	Nicholas Simmons	44
C1012	Bruce Butler	36
C1013	Raymond Alexander	49
C1014	Jason Duncan	44
C1015	Phillip Peters	41

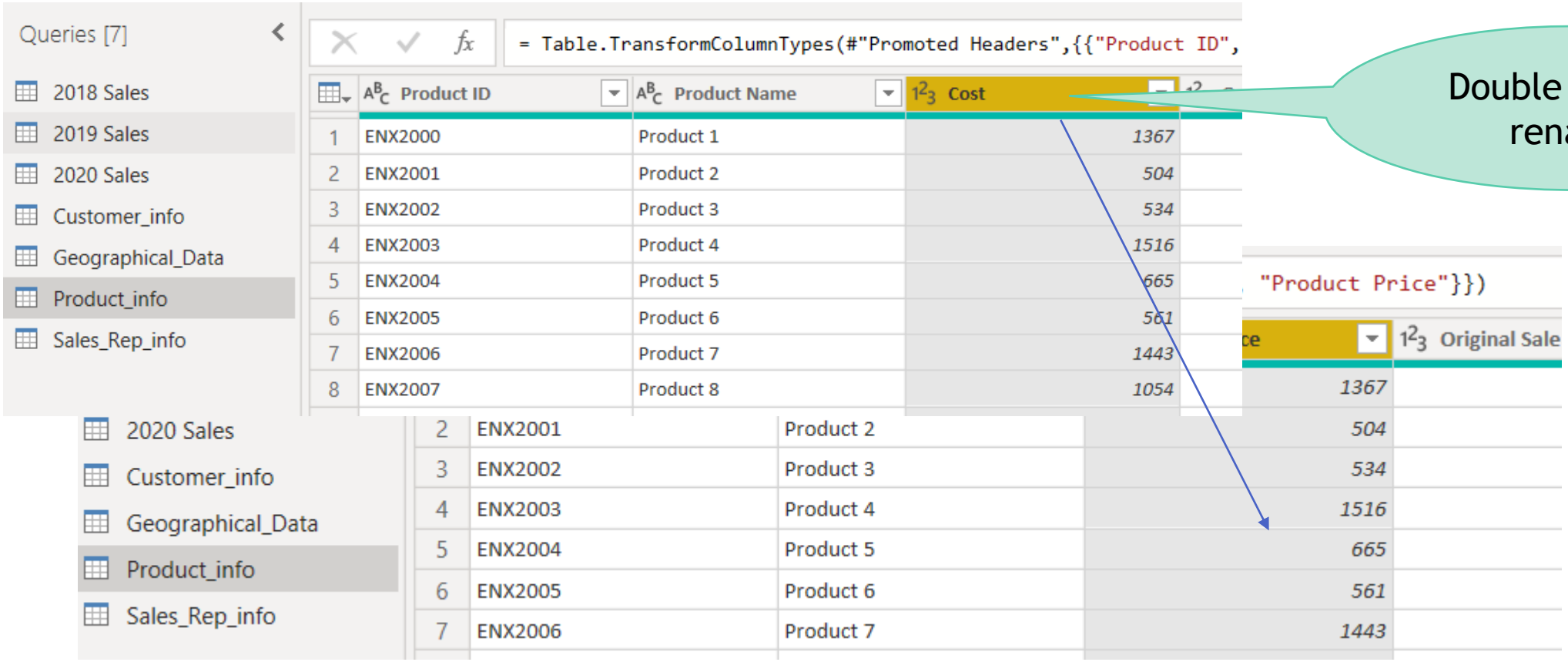
E-commerce Product Sales Analysis

Step-3: Transform the data

Transformations

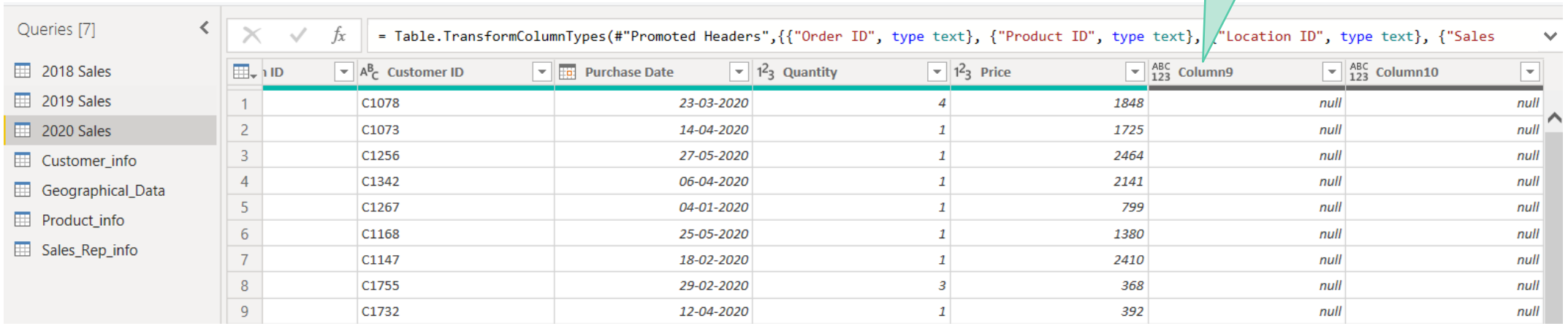
- Transform the data

Rename a column



Remove columns- Sales

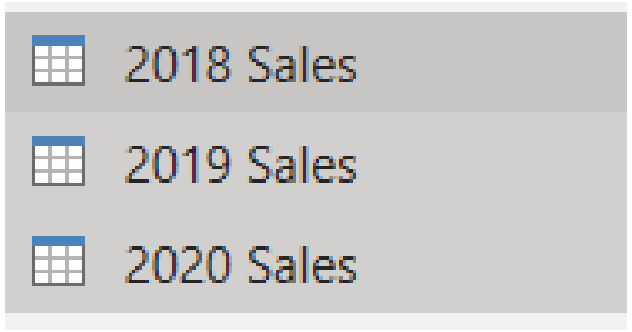
Remove the columns

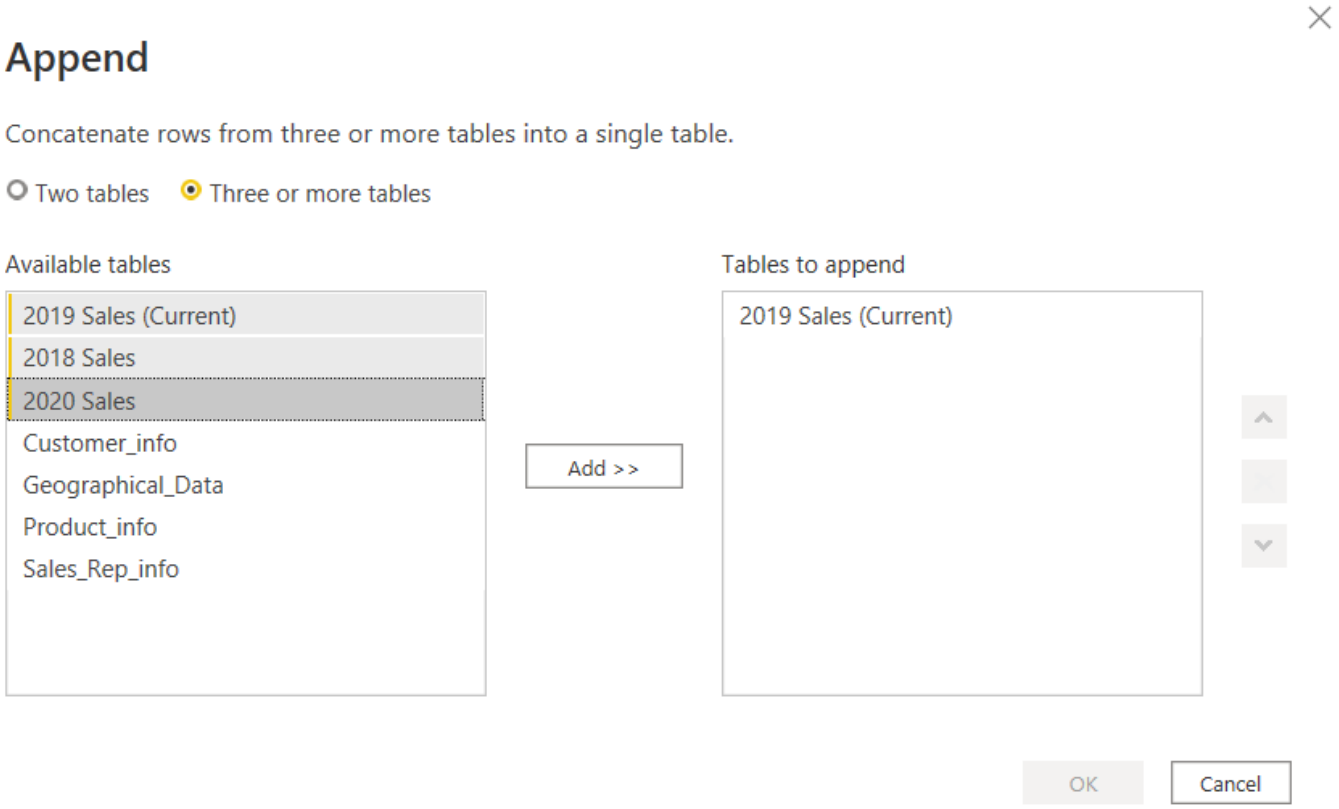


	ID	Customer ID	Purchase Date	Quantity	Price	Column9	Column10
1		C1078	23-03-2020	4	1848	null	null
2		C1073	14-04-2020	1	1725	null	null
3		C1256	27-05-2020	1	2464	null	null
4		C1342	06-04-2020	1	2141	null	null
5		C1267	04-01-2020	1	799	null	null
6		C1168	25-05-2020	1	1380	null	null
7		C1147	18-02-2020	1	2410	null	null
8		C1755	29-02-2020	3	368	null	null
9		C1732	12-04-2020	1	392	null	null

Appending the tables

Home >Append Queries





Create StateCode-1 from State

State	State_Code1	
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C
California	CALI	C

Duplicated Column

Uppercased Text

Split Column by Position

Changed Type1

Removed Columns1

Renamed Columns

Follow these steps

See all the steps in power query editor

- Home>> Advanced Editor

Geographical_Data

Display Options

```
let
    Source = Excel.Workbook(File.Contents("D:\Google Drive\Training\Job_Courses\PowerBI\Class_Material\PowerBI_Sessions_Data\Ecom_Product_Sale
    Geographical_Data_Sheet = Source{[Item="Geographical_Data",Kind="Sheet"]}[Data],
    #"Promoted Headers" = Table.PromoteHeaders(Geographical_Data_Sheet, [PromoteAllScalars=true]),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"Location ID", type text}, {"Name", type text}, {"County", type text},
    #"Removed Columns" = Table.RemoveColumns(#"Changed Type",{"Area Code", "Land Area", "Water Area", "Time Zone"}),
    #"Duplicated Column" = Table.DuplicateColumn(#"Removed Columns", "State", "State - Copy"),
    #"Uppercased Text" = Table.TransformColumns(#"Duplicated Column",{{"State - Copy", Text.Upper, type text}}),
    #"Split Column by Position" = Table.SplitColumn(#"Uppercased Text", "State - Copy", Splitter.SplitTextByPositions({0, 4}, false), {"State
    #"Changed Type1" = Table.TransformColumnTypes(#"Split Column by Position",{{"State - Copy.1", type text}, {"State - Copy.2", type text}}),
    #"Removed Columns1" = Table.RemoveColumns(#"Changed Type1",{"State - Copy.2"}),
    #"Renamed Columns" = Table.RenameColumns(#"Removed Columns1",{{"State - Copy.1", "State_Code1"}}),
    #"Reordered Columns" = Table.ReorderColumns(#"Renamed Columns",{"Location ID", "Name", "County", "State Code", "State", "State_Code1", "Ty
in
    #"Reordered Columns"
```

Create a new table with Just dates

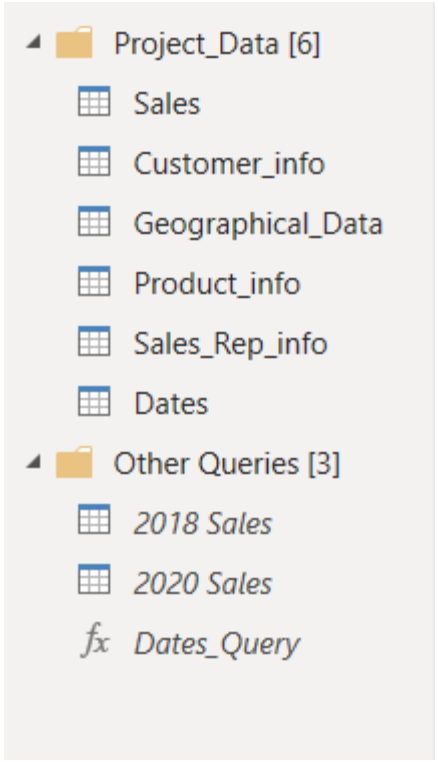
- This table will be useful for us later on
- This standard table will be helpful in future visuals and dashboards
- Most of the business analysts create this table

- Copy the date table creation M-code from Power BI Date Table code.txt file >> Home >>Advanced Editor >> Enter
- Enter the dates from 01/01/2018 to 31/12/2021
- Rename the final table - Dates

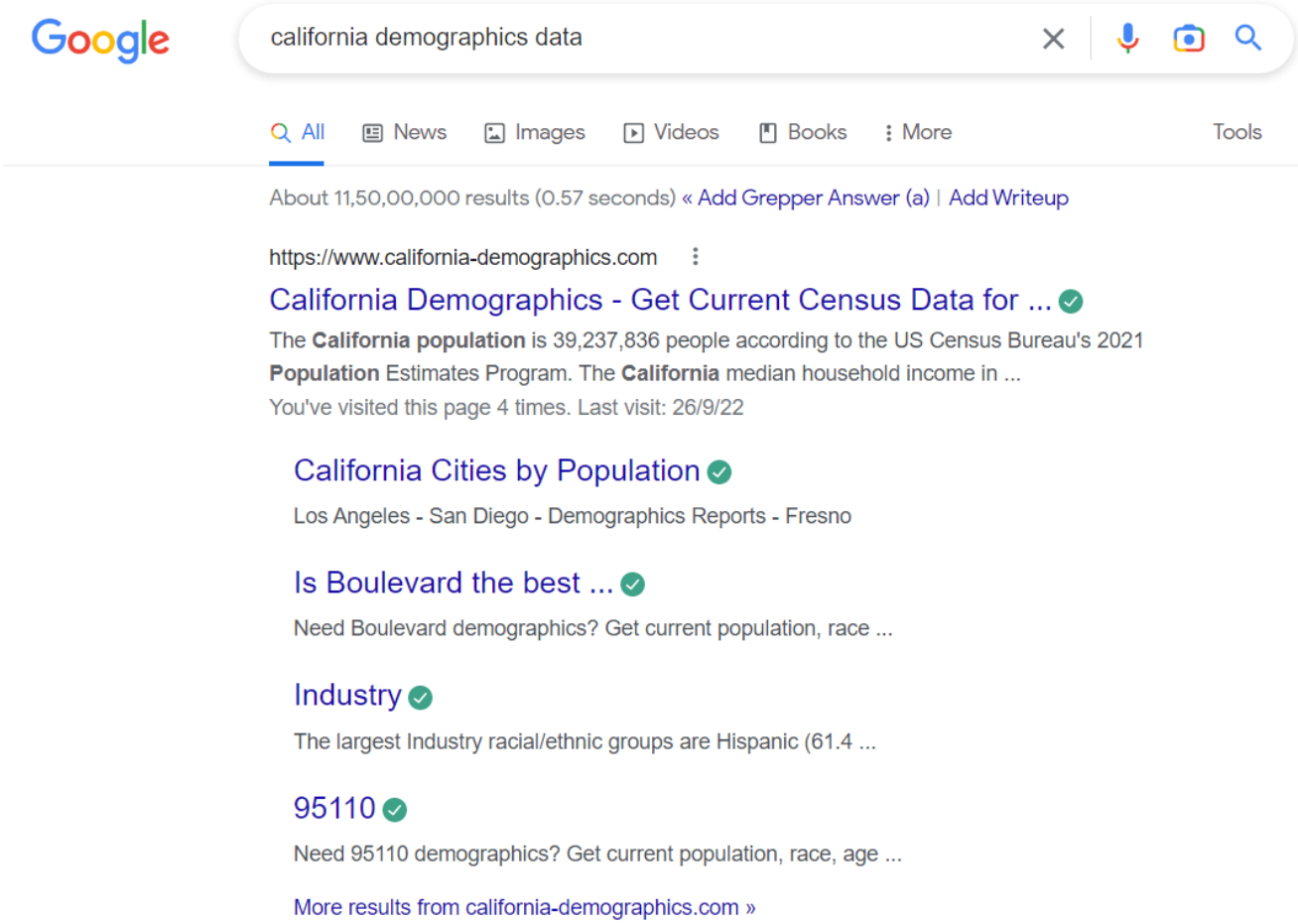
Rename Columns in the Dates Table

```
= Table.RenameColumns(Source,{{"MonthName", "Month Name"},
{"MonthInCalendar", "Month and Year"},
{"QuarterInCalendar", "Quarter and Year"}})
```

Group the tables



Bringing Data From External Sources



California Cities by Population

1	Los Angeles	3,849,297
2	San Diego	1,381,611
3	San Jose	983,489
4	San Francisco	815,201
5	Fresno	544,510
6	Sacramento	525,041
7	Long Beach	456,062
8	Oakland	433,823
9	Bakersfield	407,615
10	Anaheim	345,940

See all California Cities by Population>>

Bringing Data From External Sources

California Counties by Population

1	Los Angeles County	9,829,544
2	San Diego County	3,286,069
3	Orange County	3,167,809
4	Riverside County	2,458,395
5	San Bernardino County	2,194,710
6	Santa Clara County	1,885,508
7	Alameda County	1,648,556
8	Sacramento County	1,588,921
9	Contra Costa County	1,161,413
10	Fresno County	1,013,581

See all California Counties by Population>>

California Cities by Population

1	Los Angeles	3,849,297
2	San Diego	1,381,611
3	San Jose	983,489
4	San Francisco	815,201
5	Fresno	544,510
6	Sacramento	525,041
7	Long Beach	456,062
8	Oakland	433,823
9	Bakersfield	407,615
10	Anaheim	345,940

See all California Cities by Population>>

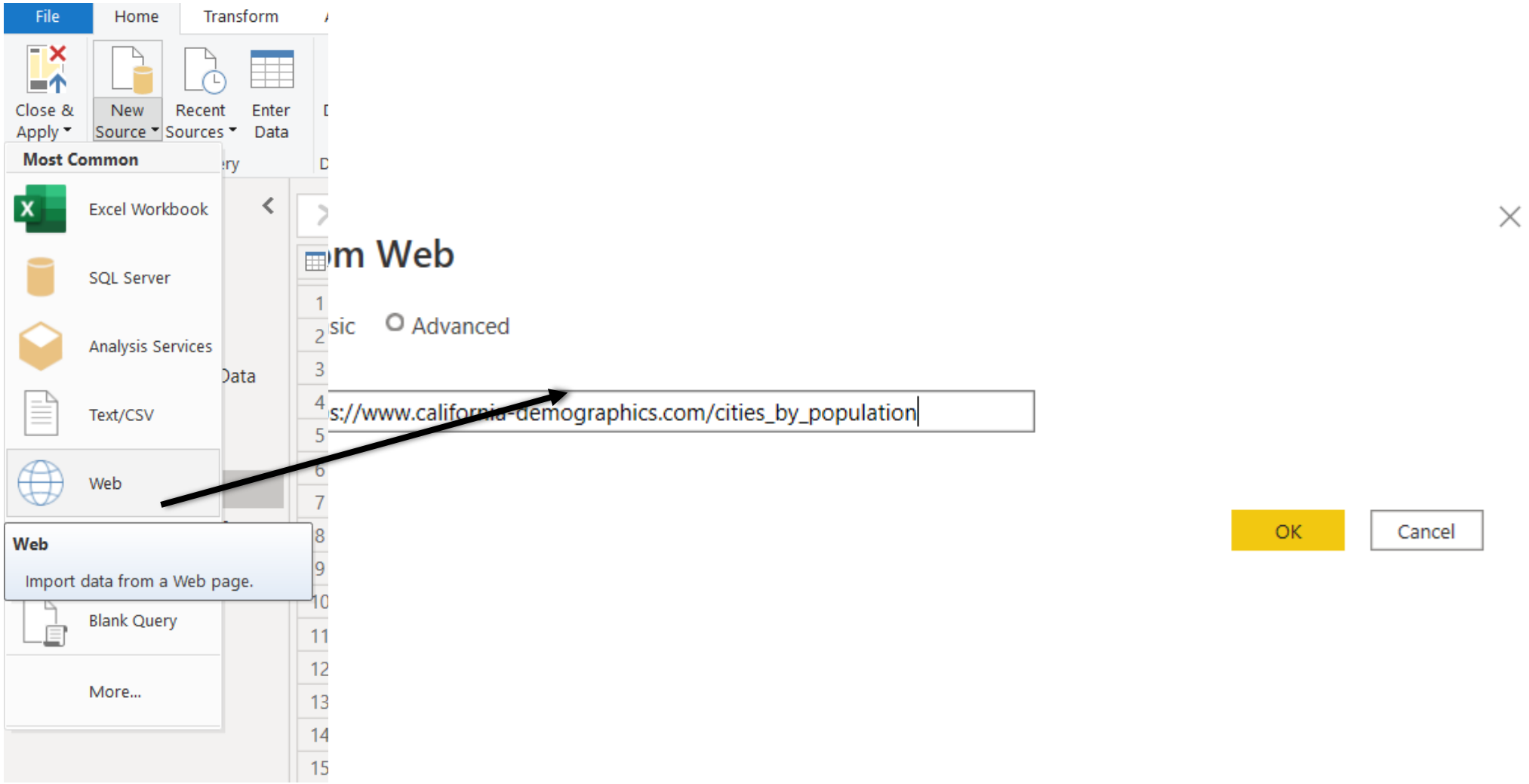
California Zip Codes by Population

1	90011	110,750
2	90650	104,524
3	91331	103,683
4	90201	100,512
5	90044	99,980
6	92335	99,242
7	92336	98,731
8	94565	97,249
9	90250	95,115
10	90805	95,094

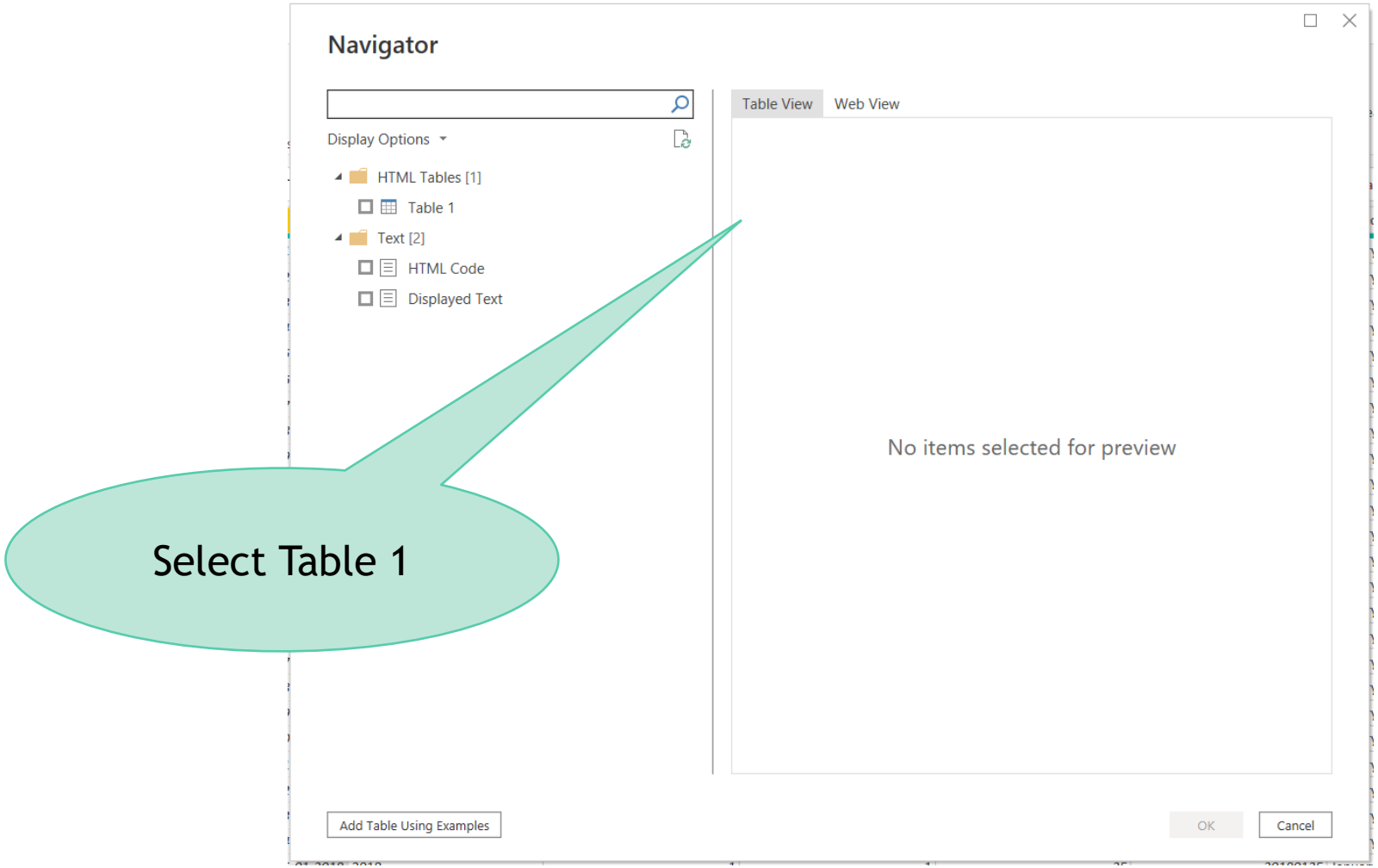
See all California Zip Codes by Population>>

Click on cities population

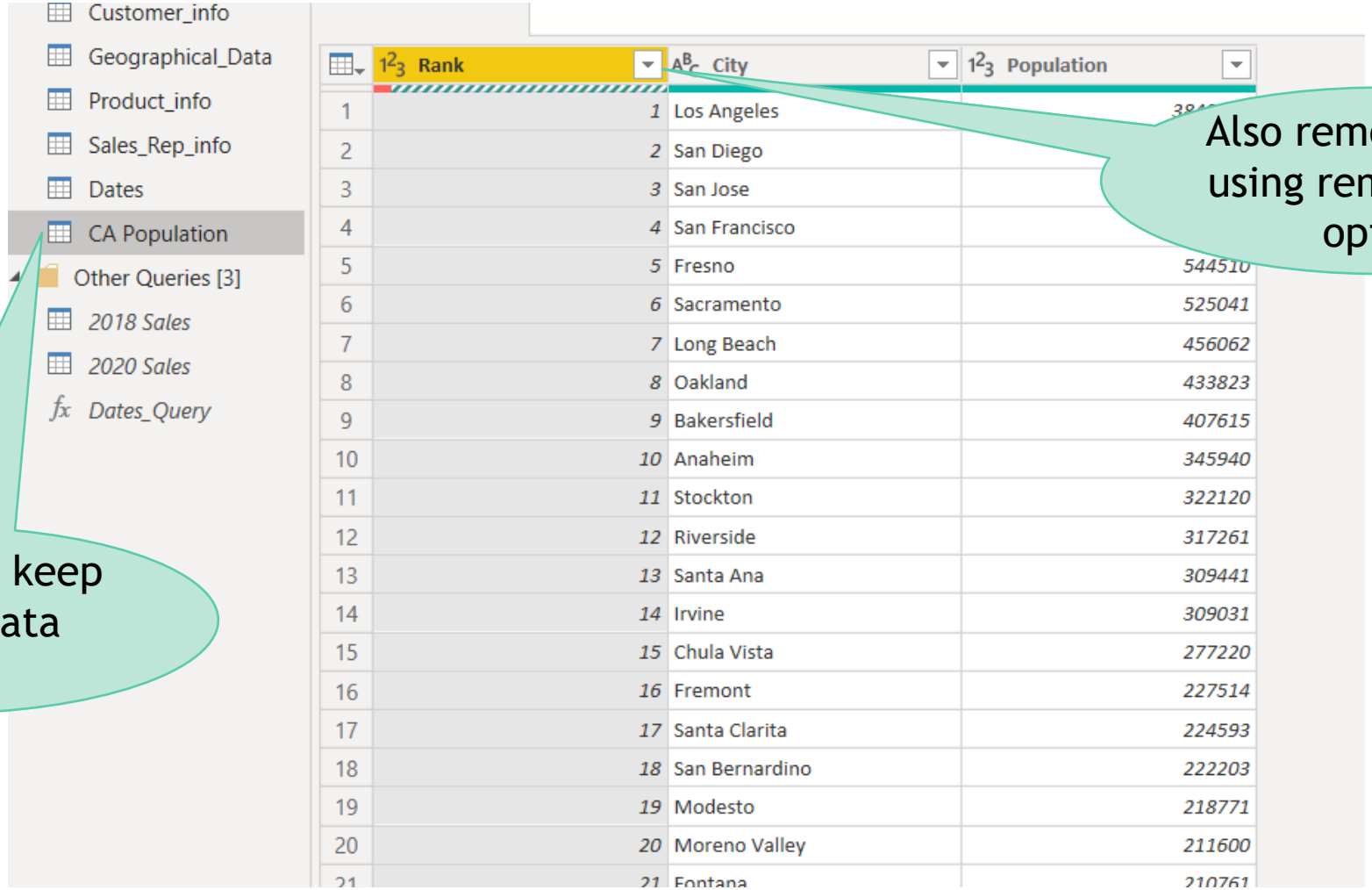
Copy the data from website



Copy the data from website



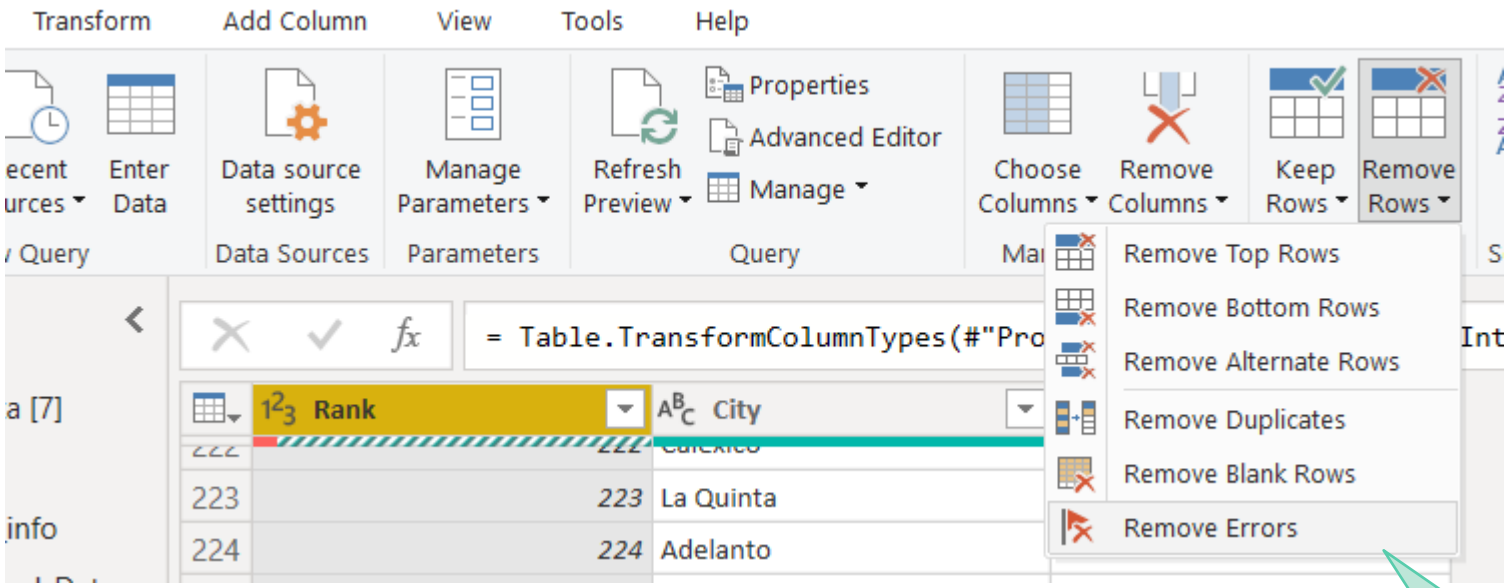
Copy the data from website



Also remove errors using remove rows option

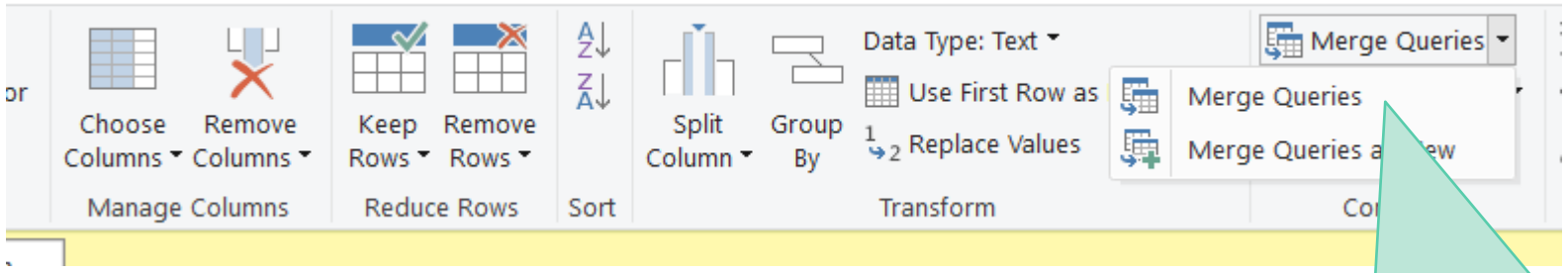
Rename and keep it in the data group

Copy the data from website



Select the column >> remove rows

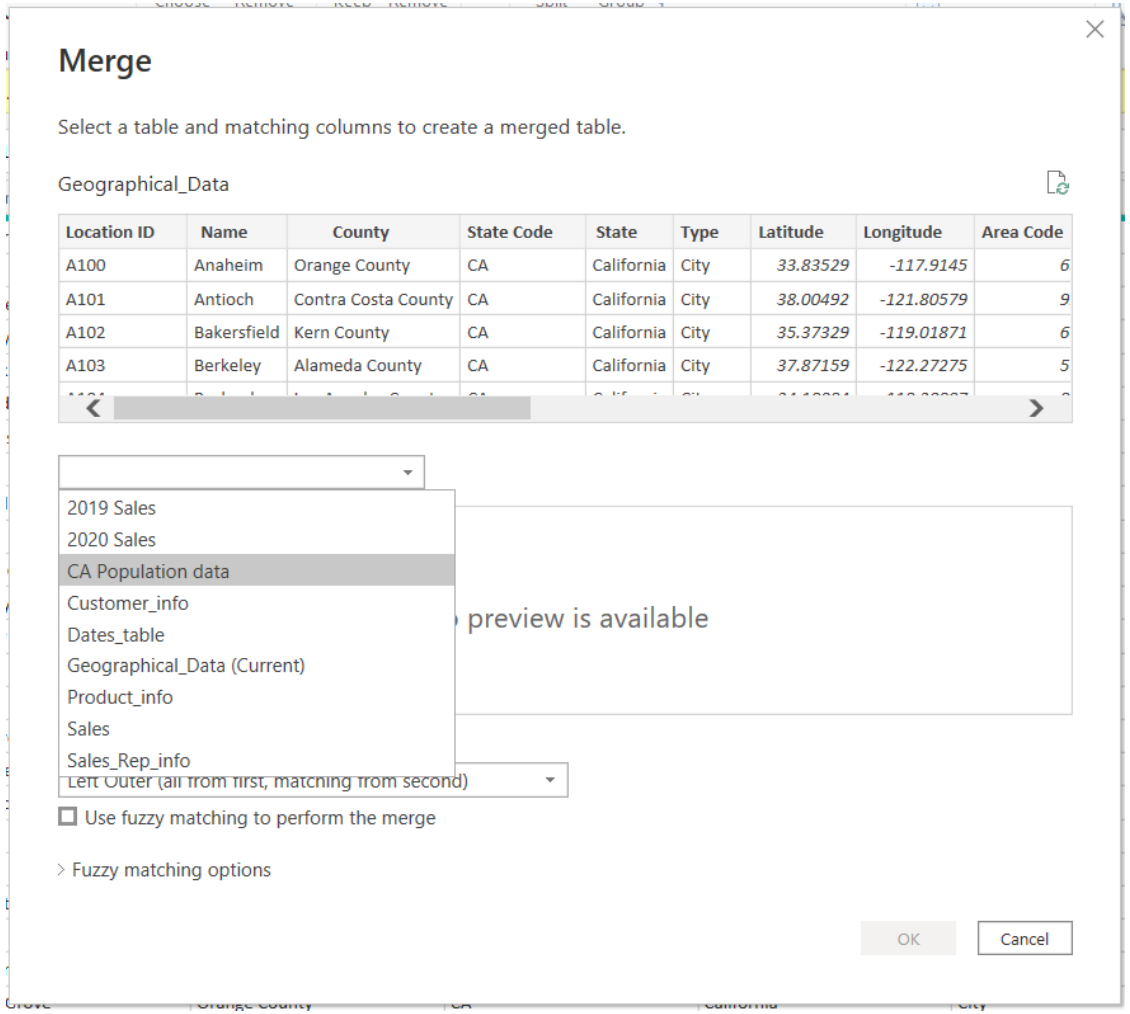
Merge Queries



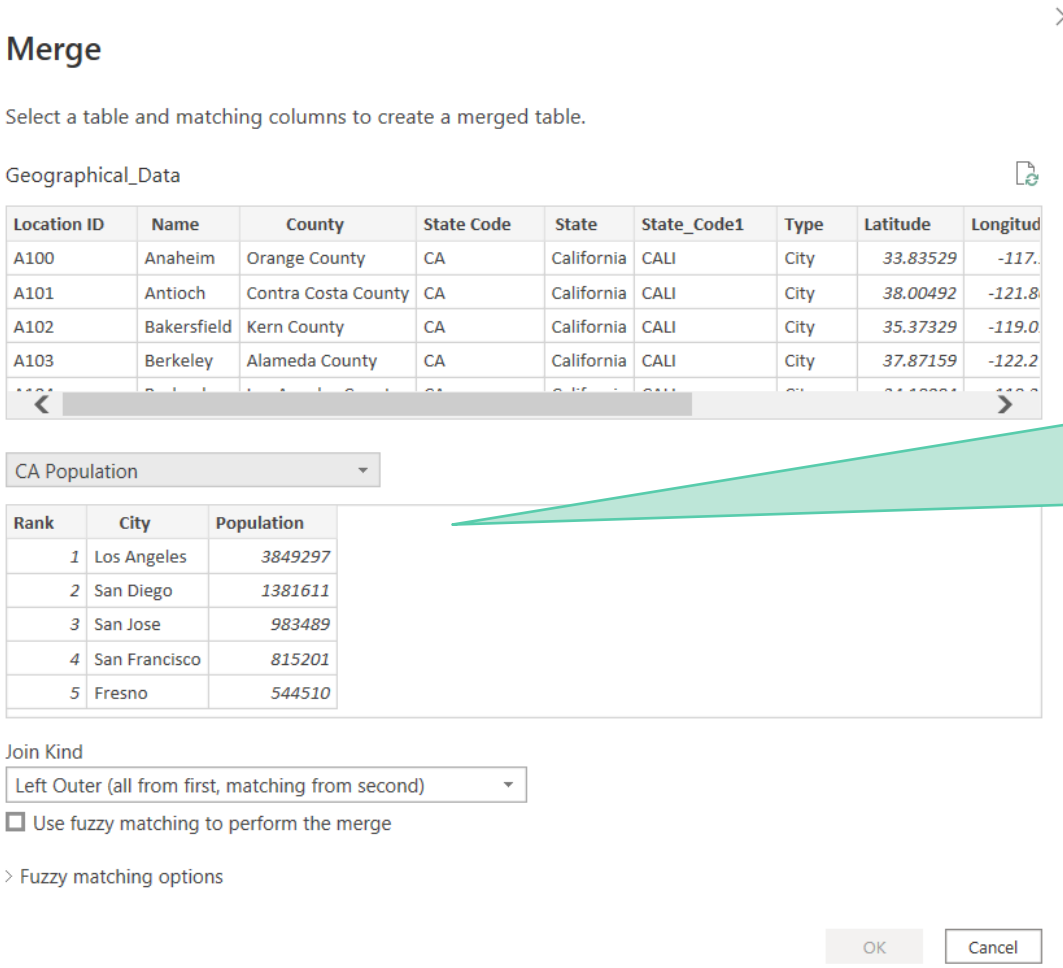
We would like to merge geographical data.

Select Geographical table >> Home >> Merge Queries

Merge Queries

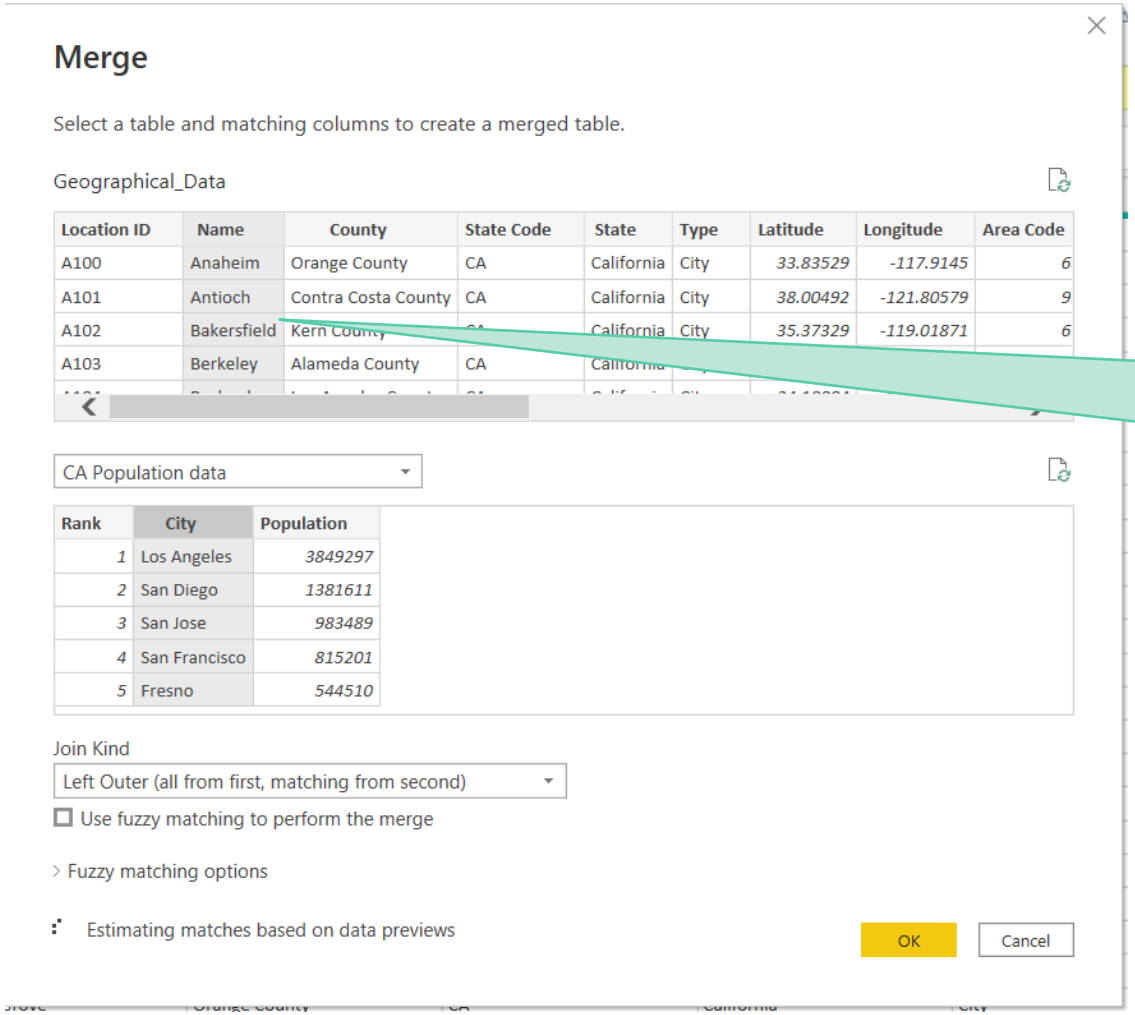


Merge Queries



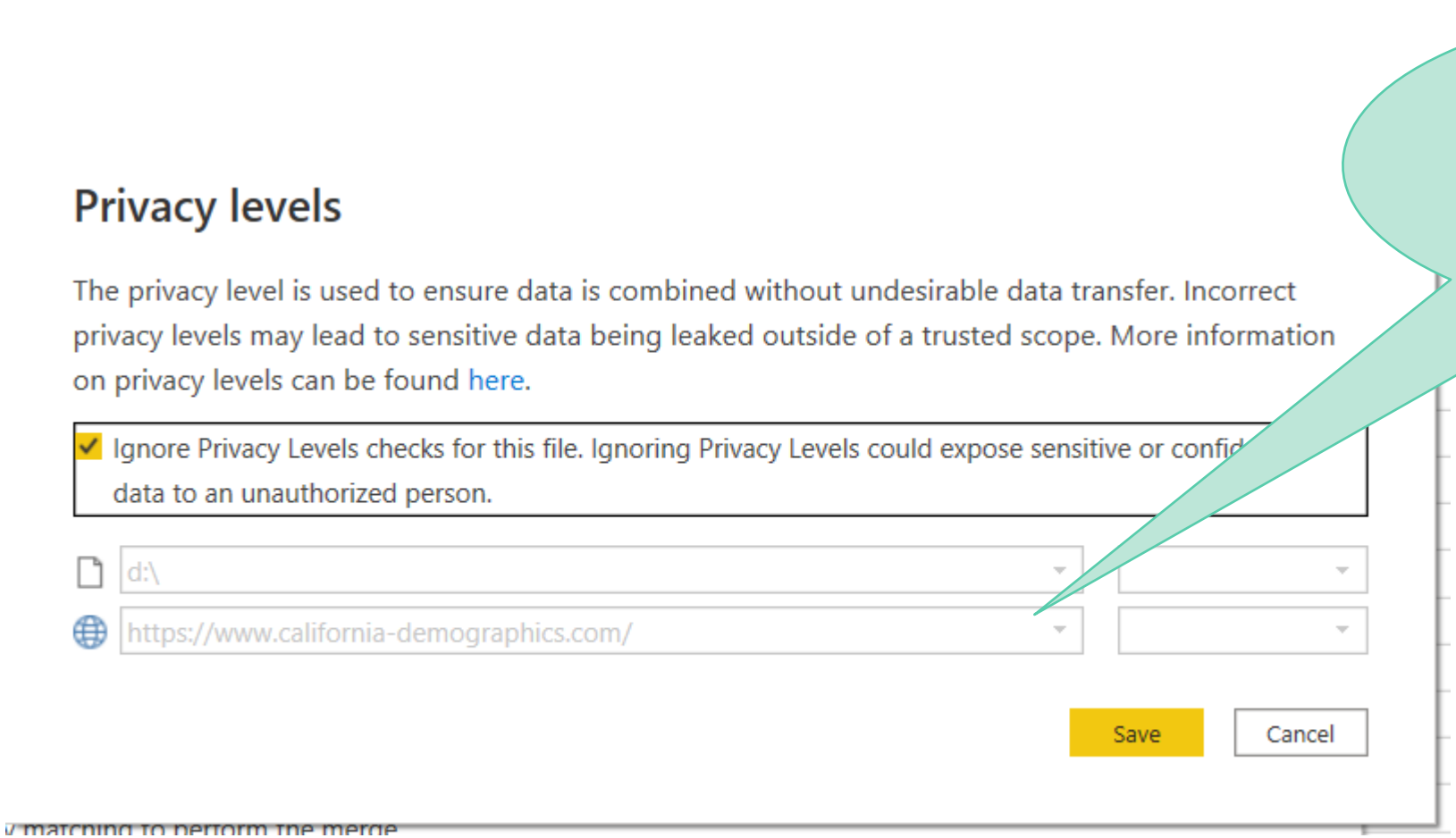
Geographical Data >> Merge Queries >> Select the join

Merge Queries



Select fields to join upon

Merge Queries



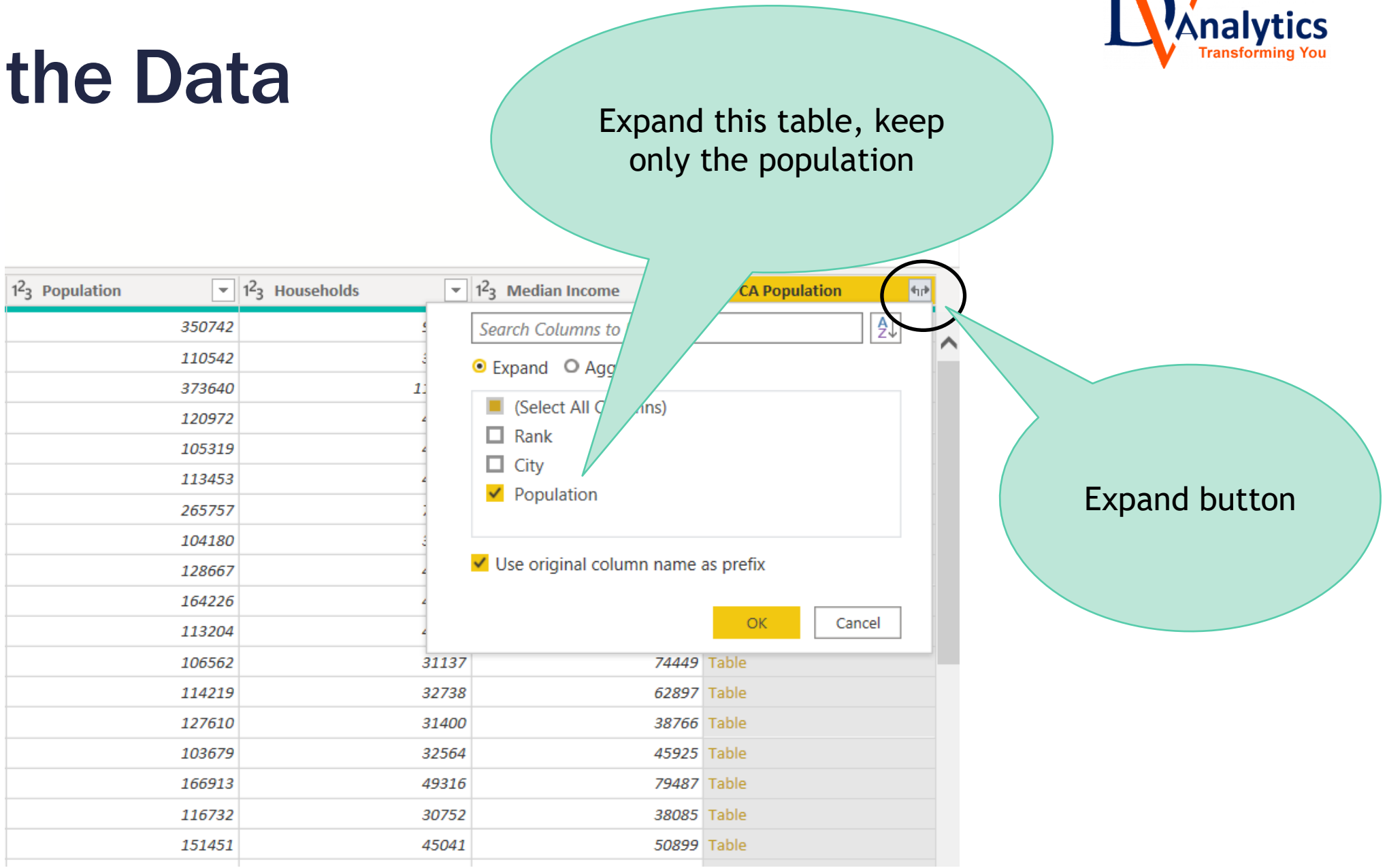
Appears only once, for the security sake

Attach the Data

Expand this table, keep only the population

	1.2 Latitude	1.2 Longitude	123 Population	123 Households	123 Median Income	CA Population
1	33.83529	-117.9145	350742	99670	60752	Table
2	38.00492	-121.80579	110542	33718	64329	Table
3	35.37329	-119.01871	373640	112439	57095	Table
4	37.87159	-122.27275	120972	45917	66237	Table
5	34.18084	-118.30897	105319	41361	66076	Table
6	33.15809	-117.35059	113453	42791	90597	Table
7	32.64	-117.08417	265757	78066	65185	Table
8	36.82523	-119.70292	104180	34512	62666	Table
9	37.97798	-122.03107	128667	45409	68318	Table
10	33.87529	-117.56644	164226	48156	74149	Table
11	33.64113	-117.91867	113204	40908	66459	Table
12	37.70583	-122.46194	106562	31137	74449	Table

Attach the Data



Attach the Data

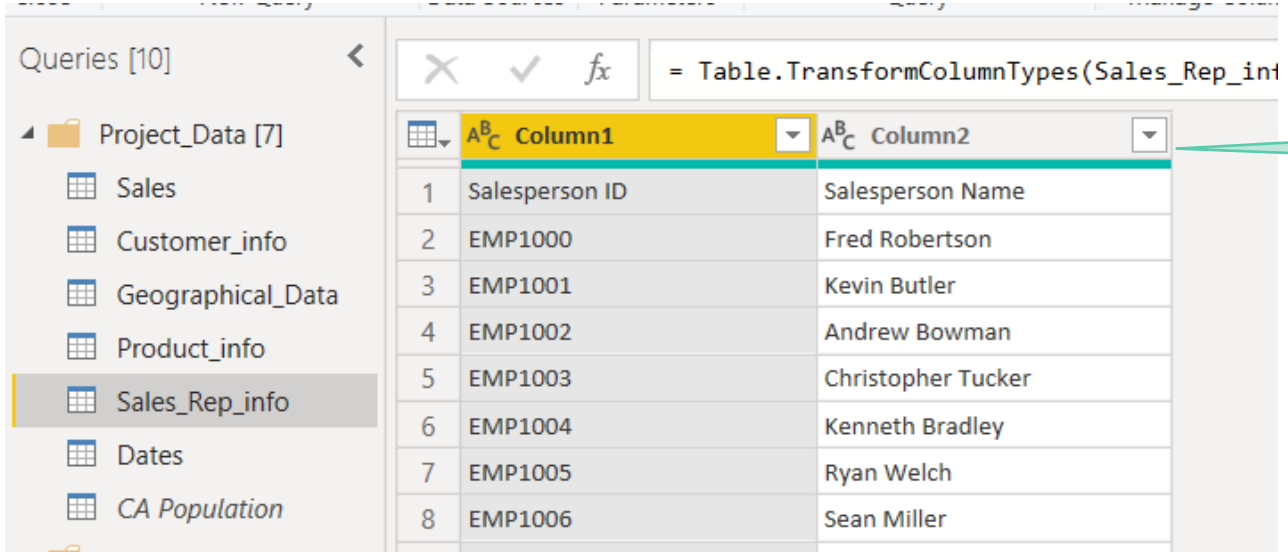
Rename the two columns

```
= Table.RenameColumns(#"Expanded CA Population data",{{"CA Population data.Population", "Updated_Population"}, {"Name", "City Name"}})
```

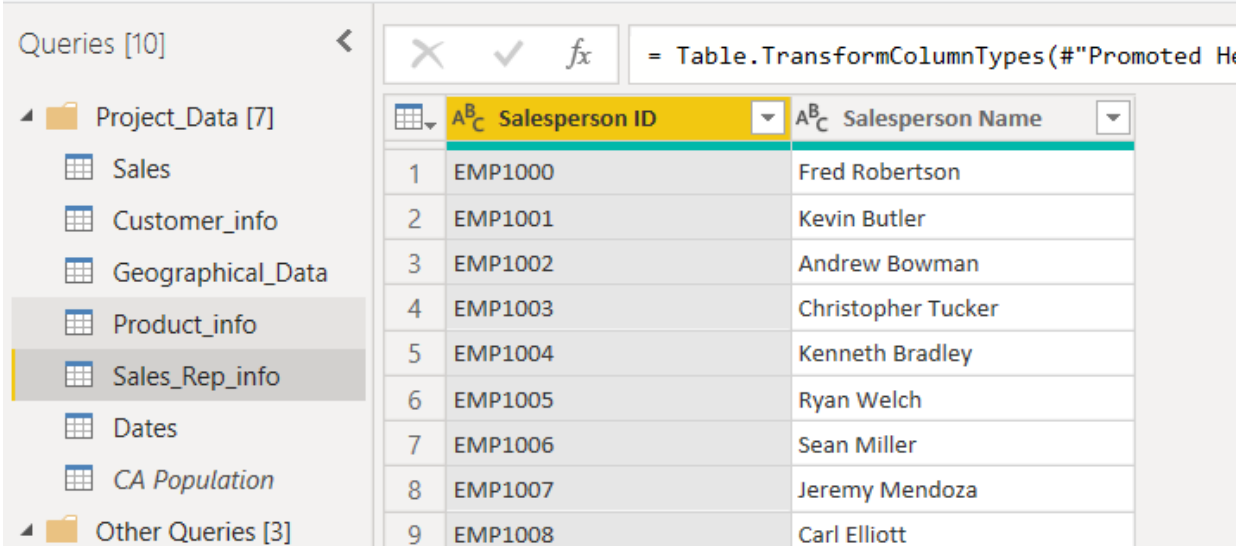
	Updated_population
60752	345940
50205	3849297
64329	114794
66116	1381611
57095	407615
84647	983489
66237	117145
81294	815201
41531	544510
90597	115302

Rename the column

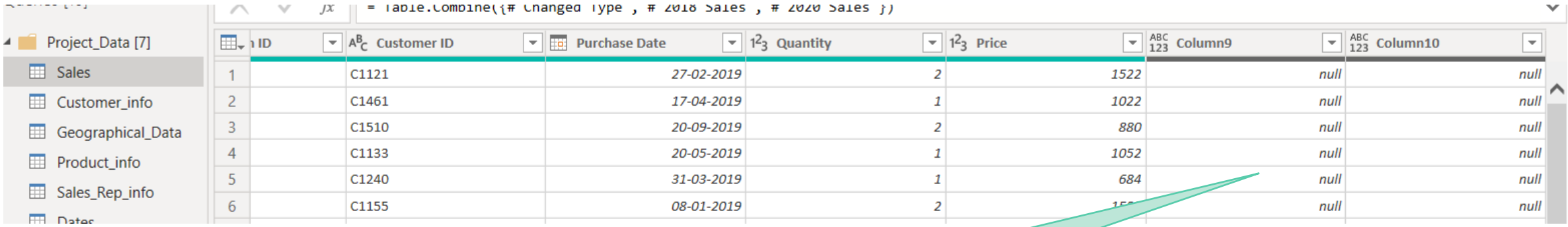
Further transformations



Use first row as headers



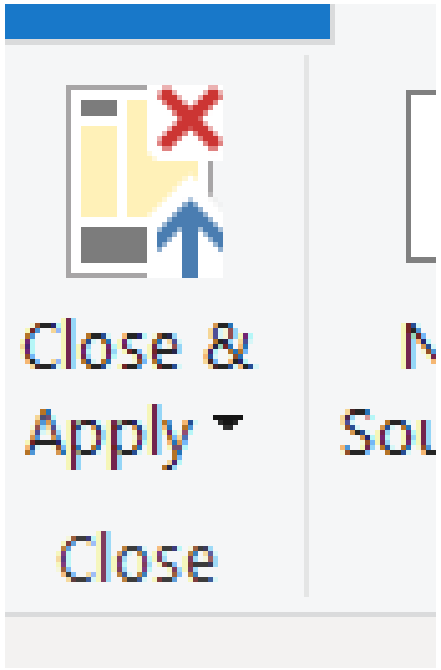
Further transformations

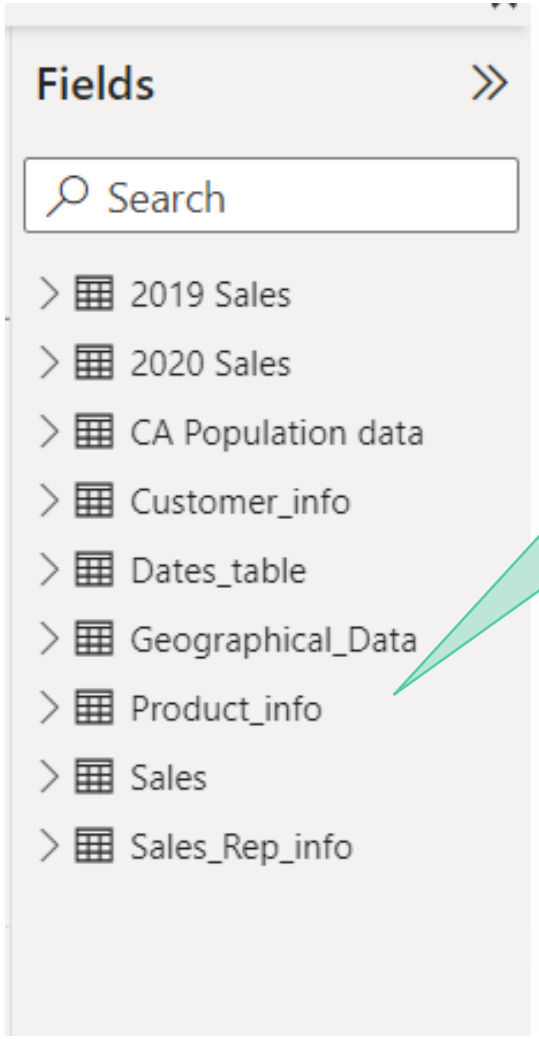


Remove last two columns from the Sales table

Close and apply

- To save all the queries until now





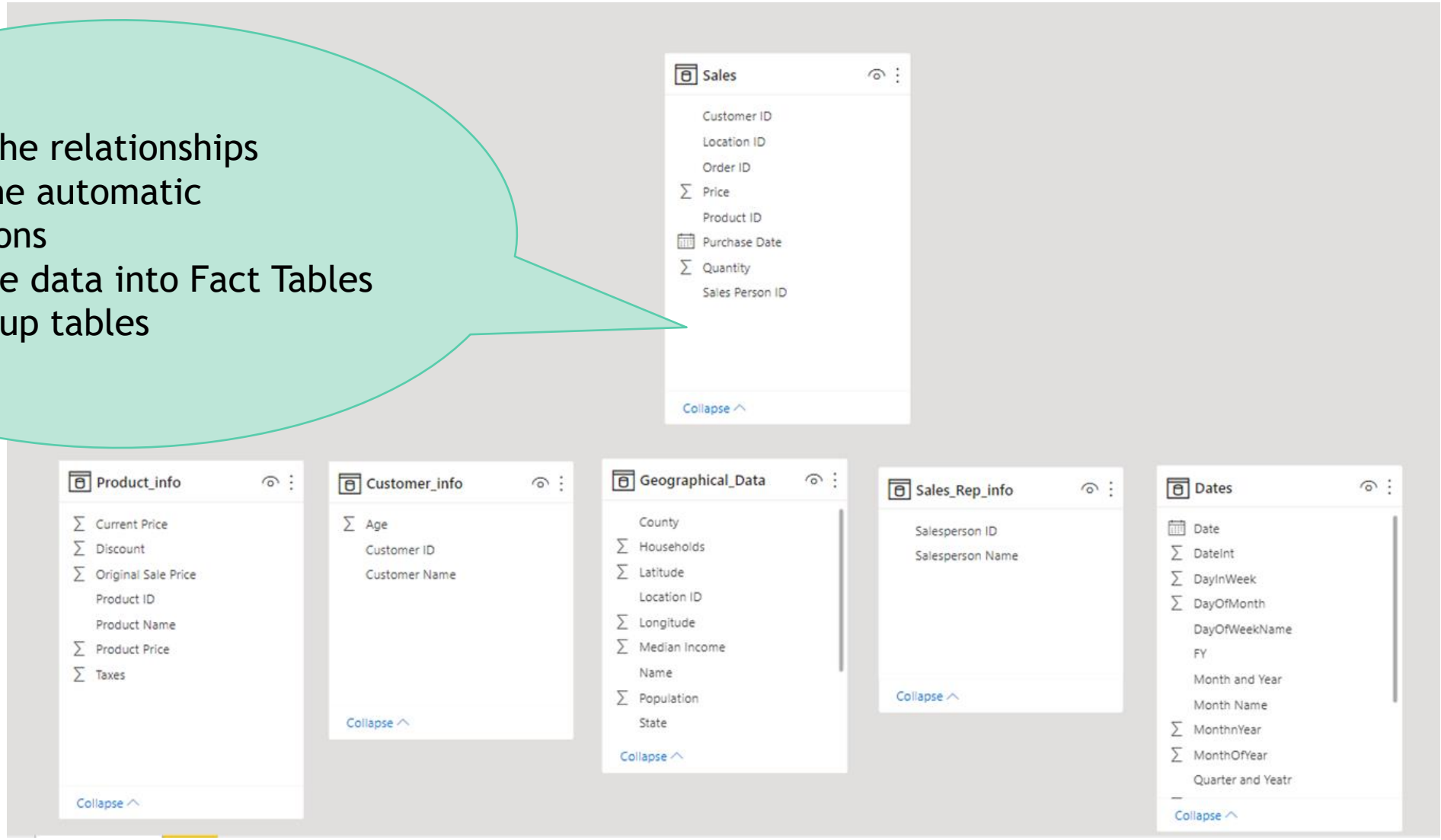
Make sure that all the tables are here, else Transform Data >> Enable load

Product Sales Case study

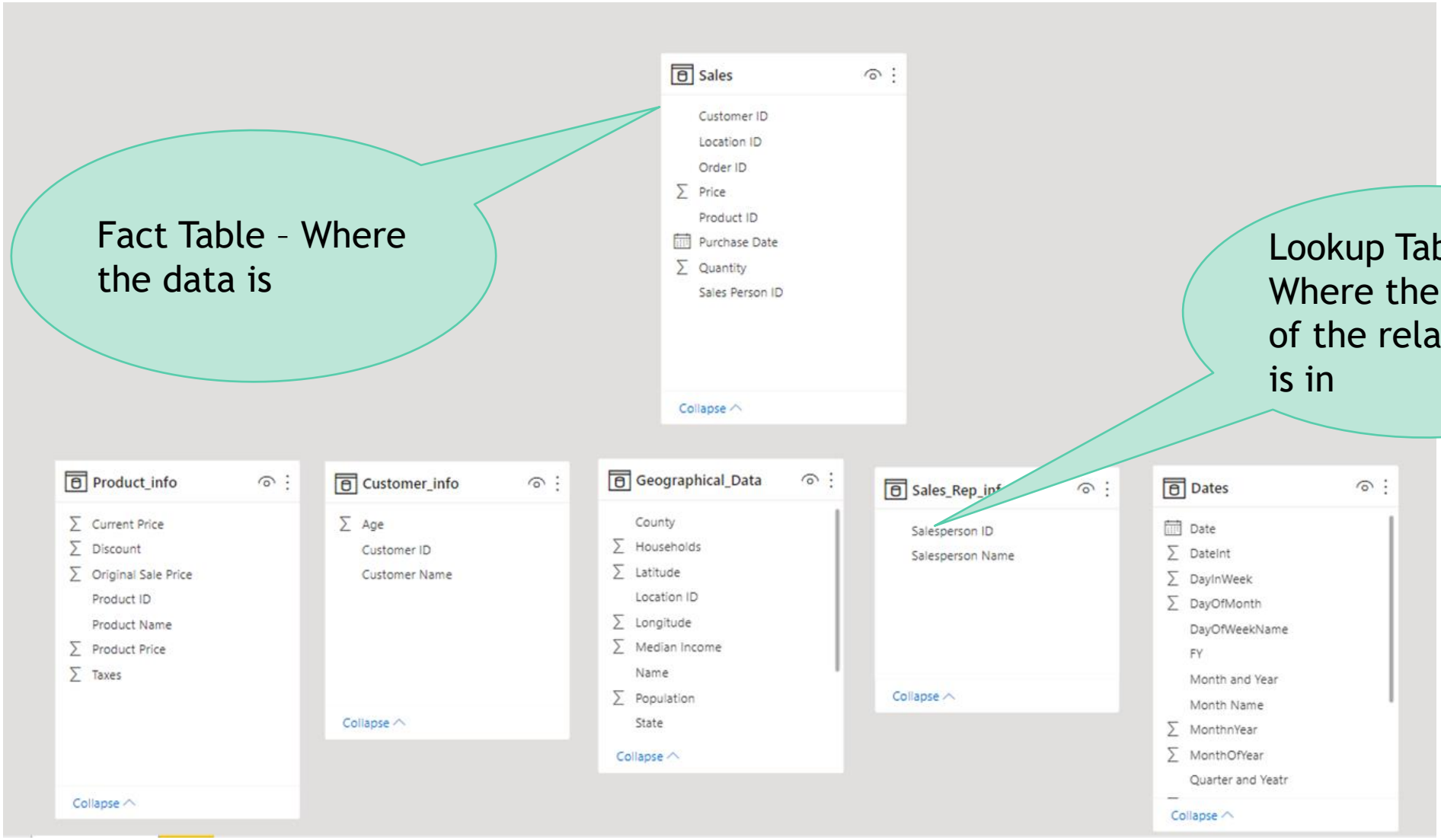
Step4: Model the data and create relationships

Model –Relationships

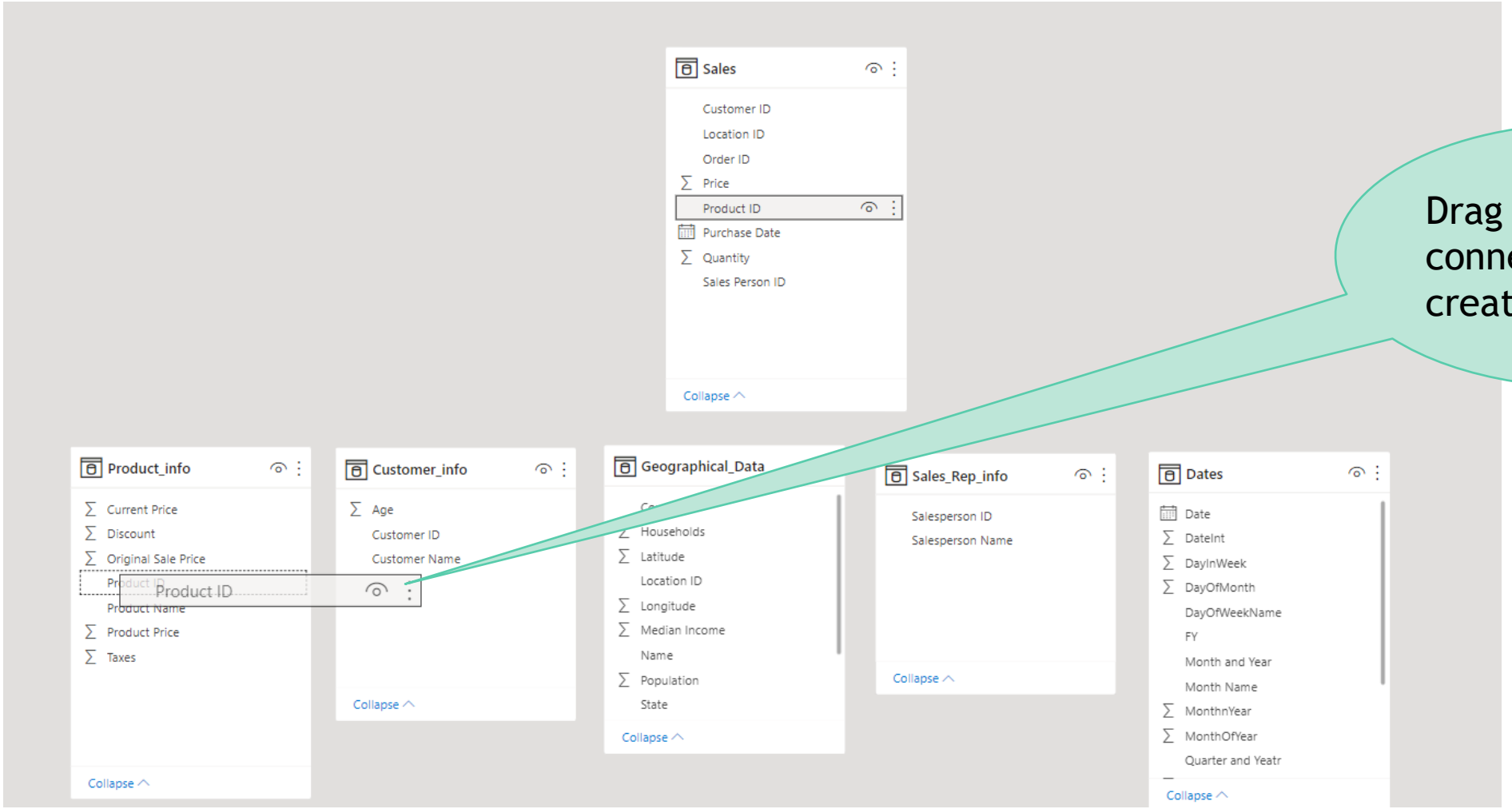
Look at the relationships
Delete the automatic connections
Divide the data into Fact Tables and Lookup tables



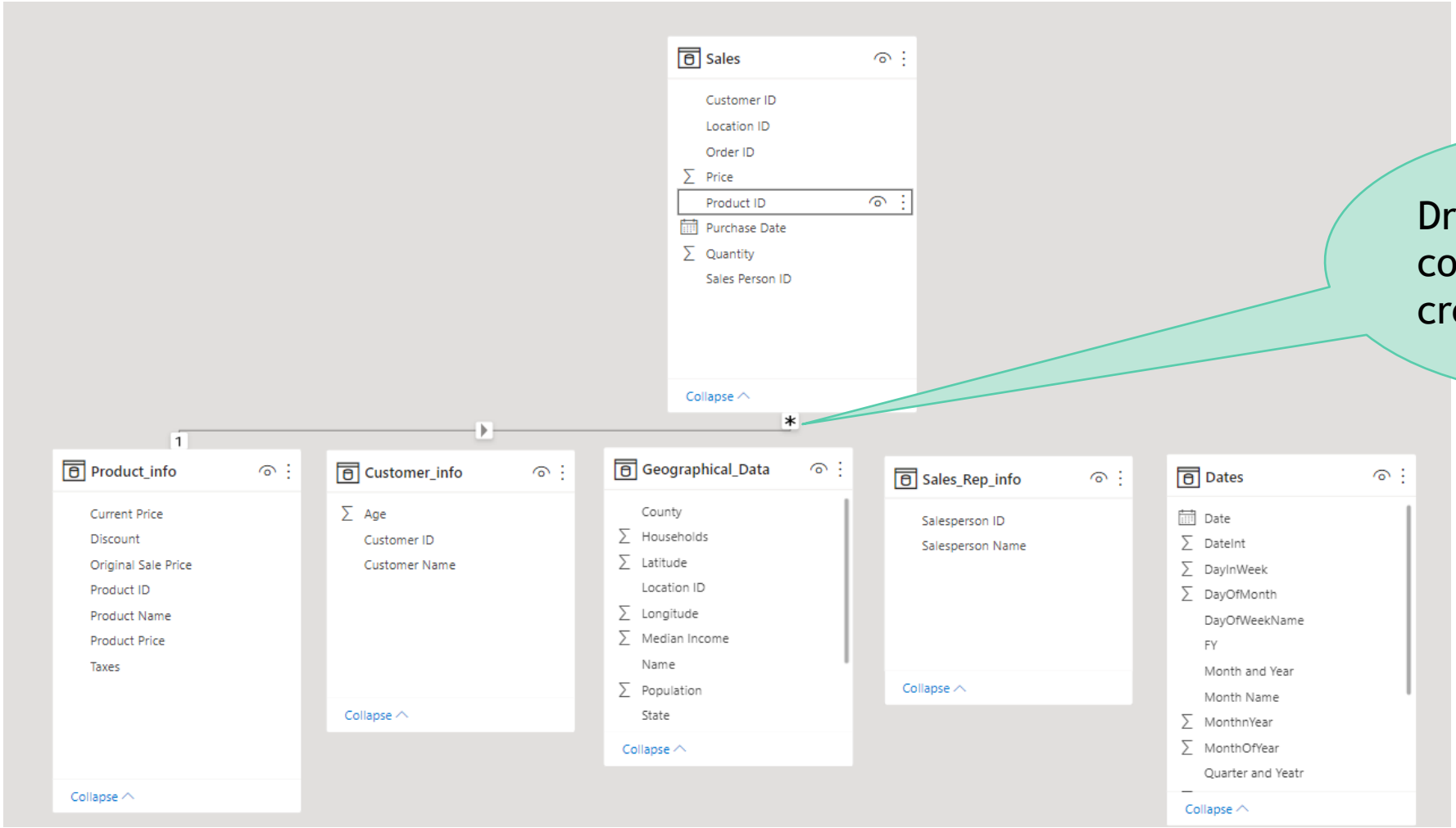
Fact Table – Lookup Table



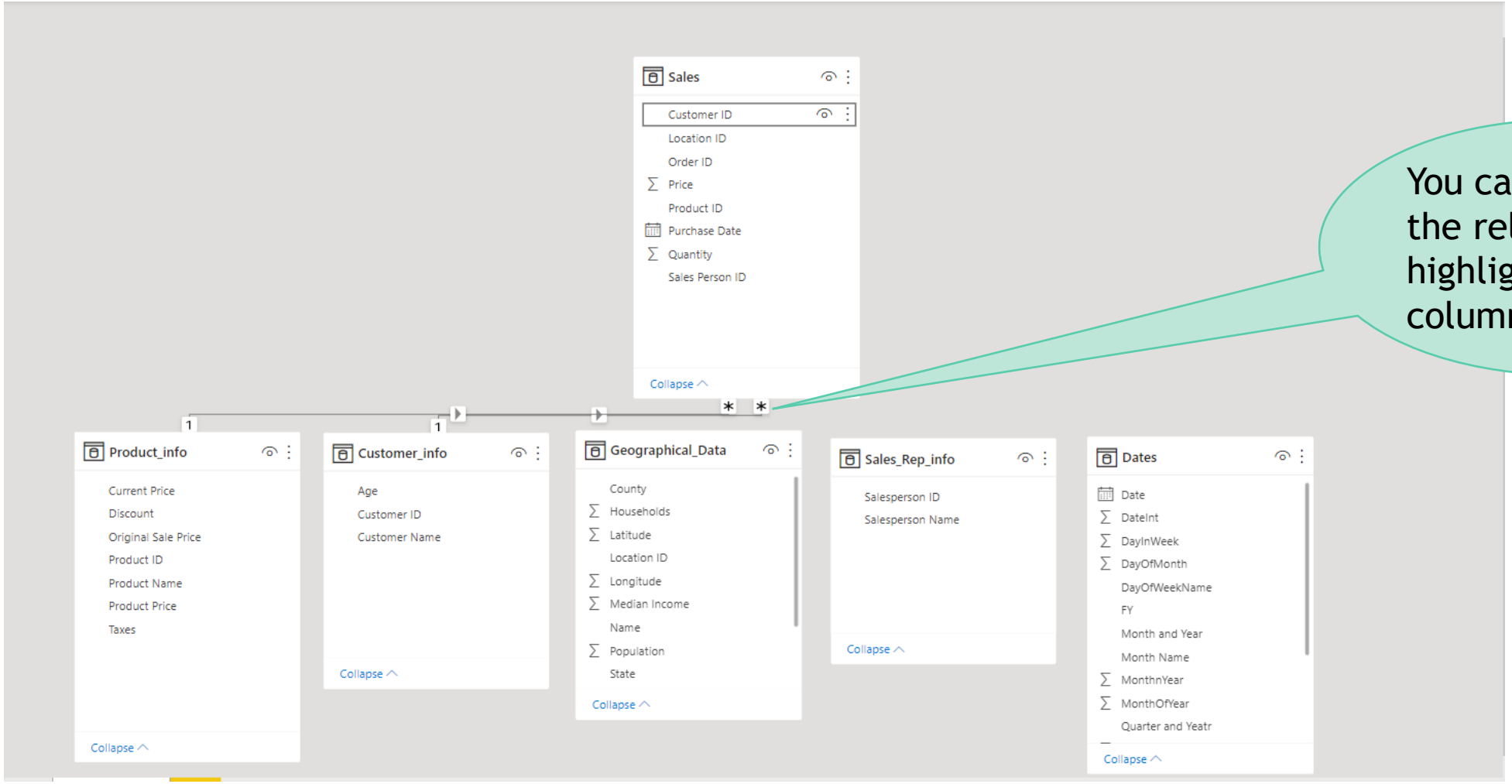
Create Connections



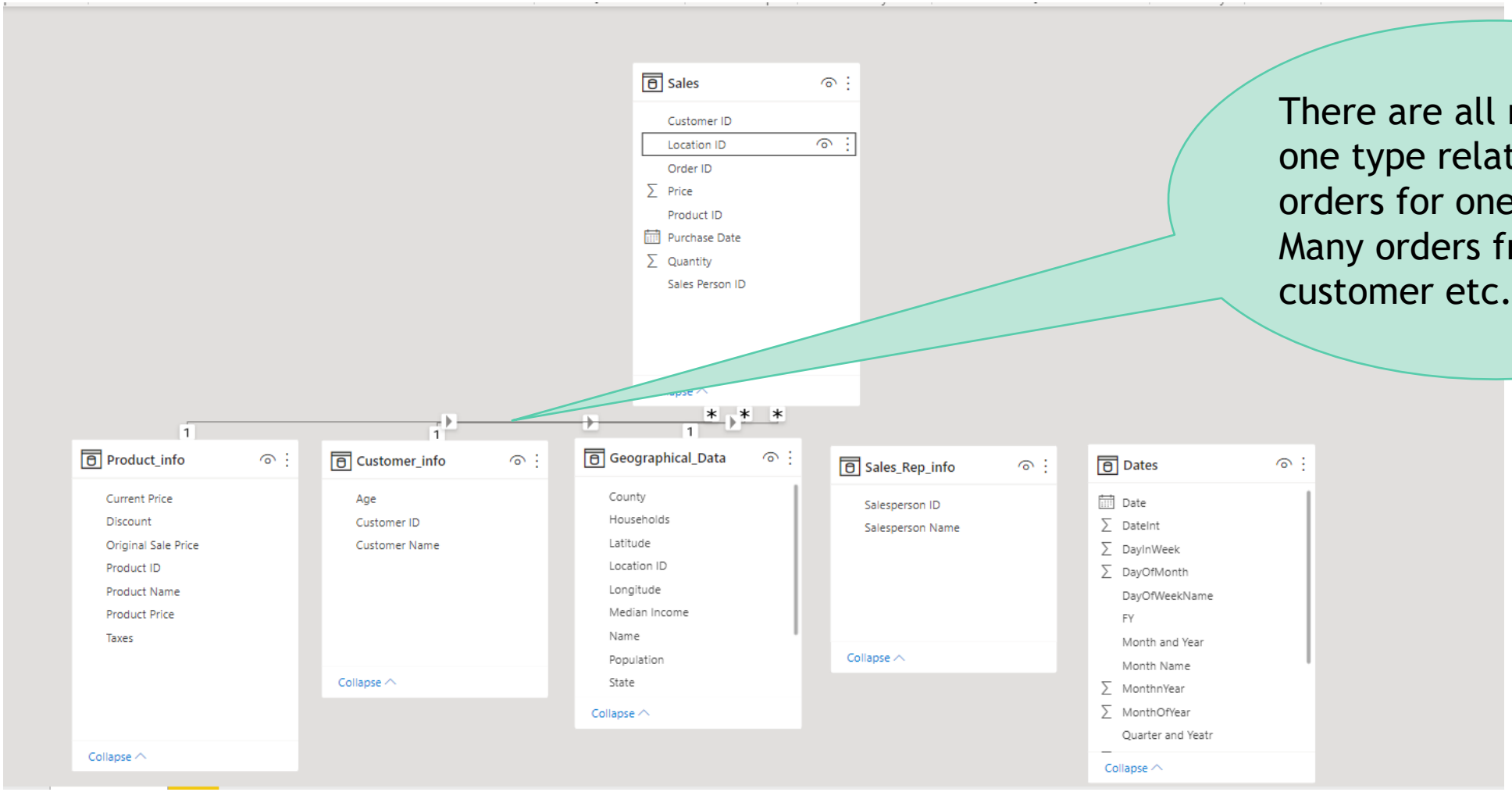
Create Connections



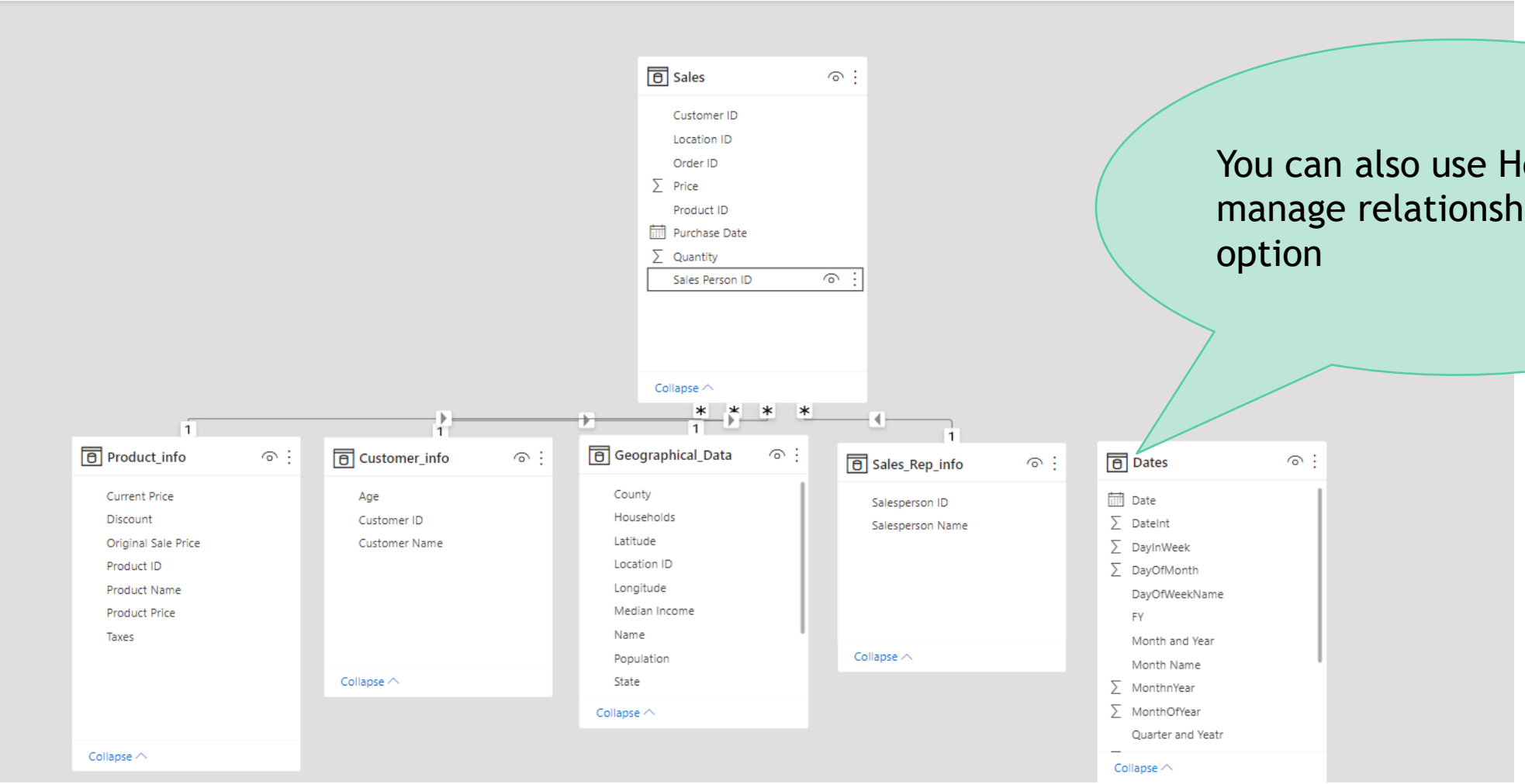
Create Connections – Other Lookup Tables



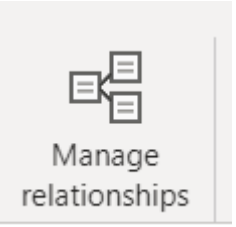
Create Connections – Other Lookup Tables

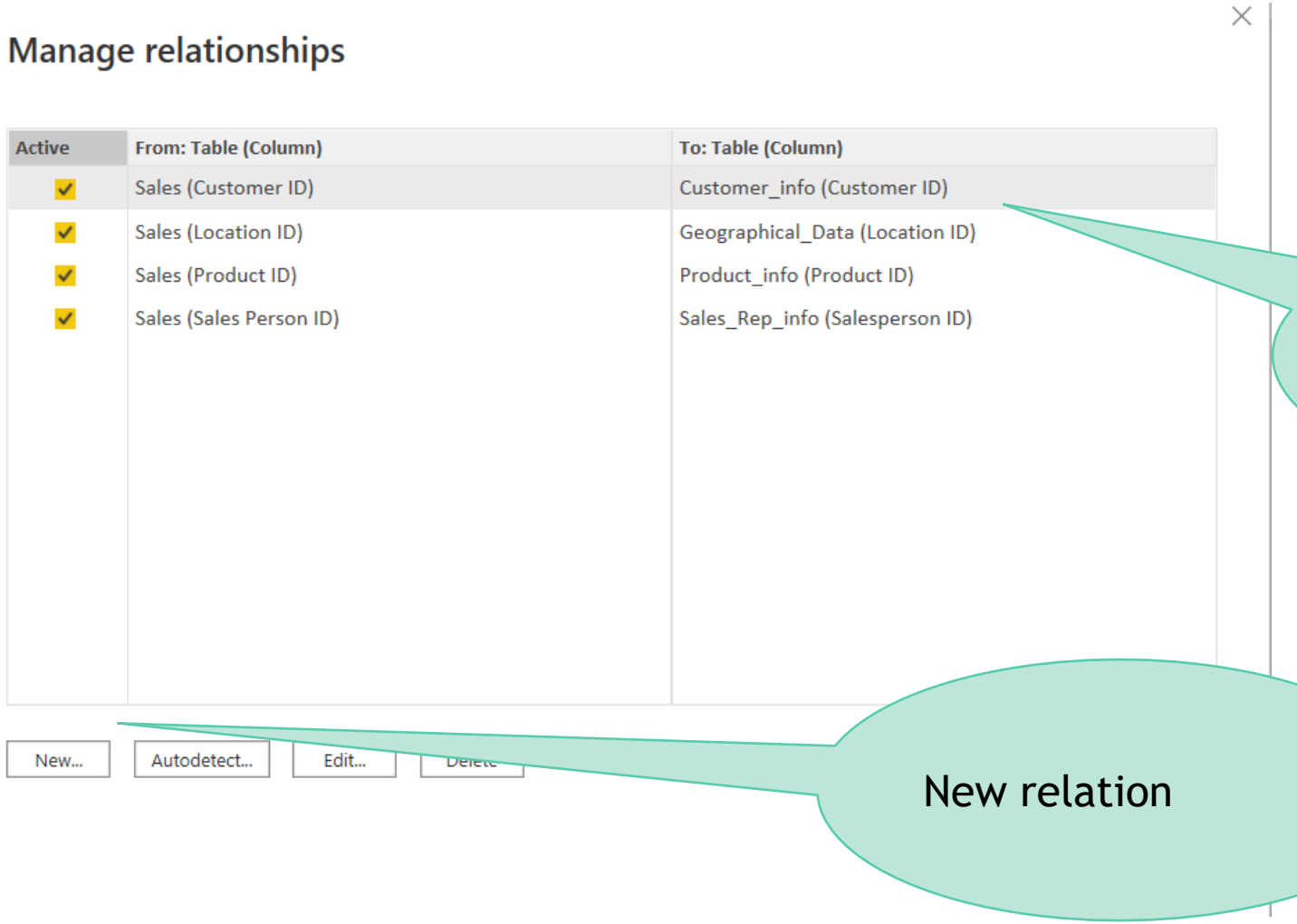


Create Connections – Other Lookup Tables

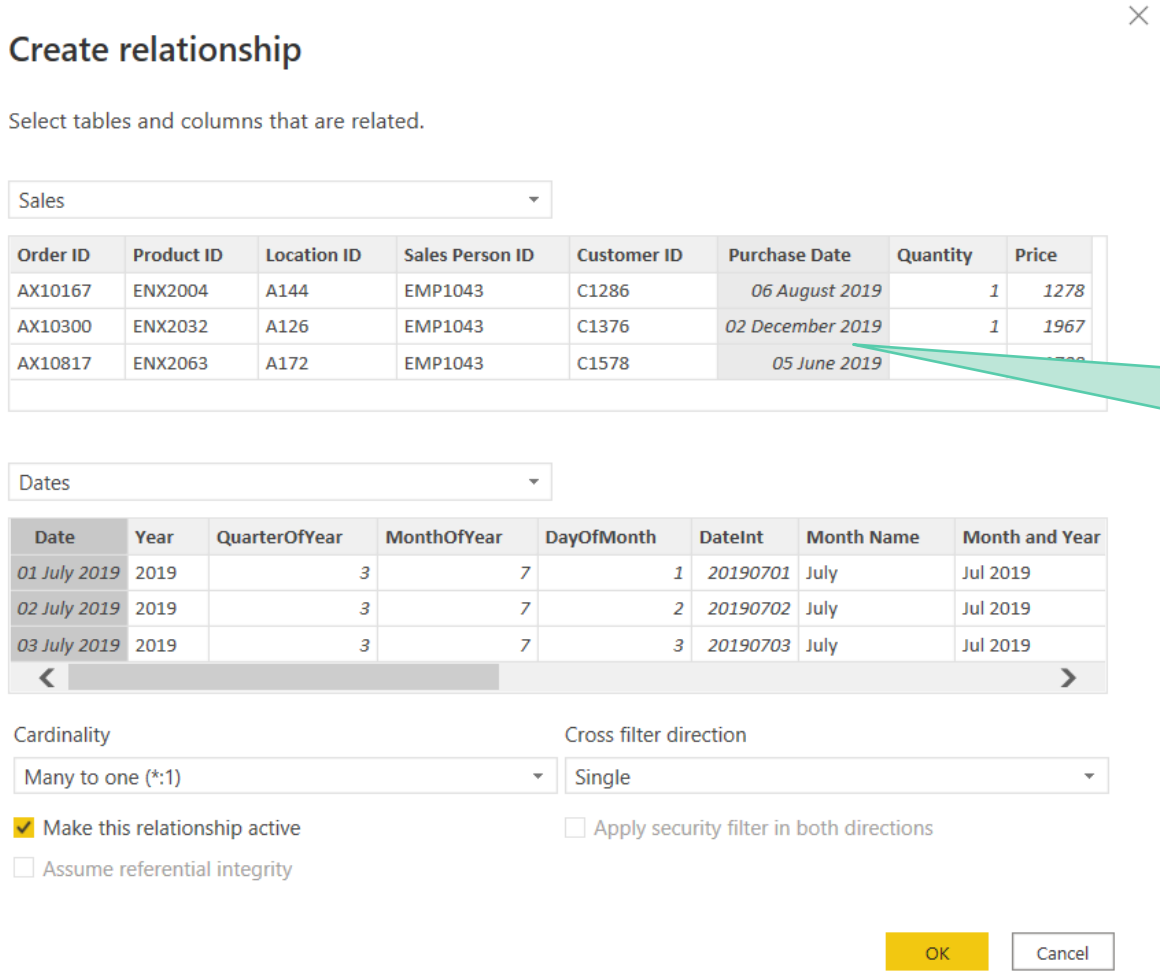


Manage table relationships



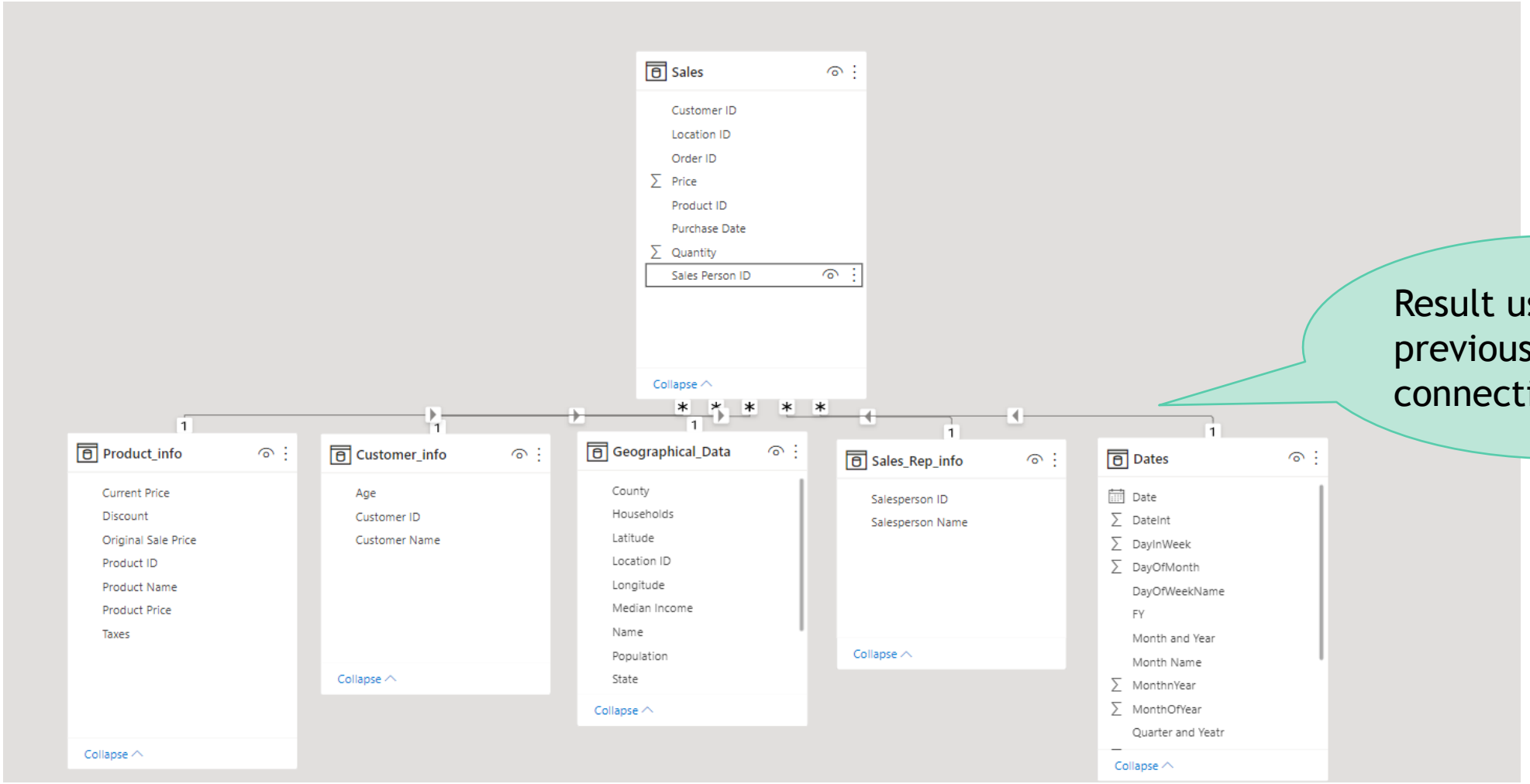


Manage table relationships



Dates column is joined based on purchase date

Manage table relationships



Next Step

DAX Deep Dive