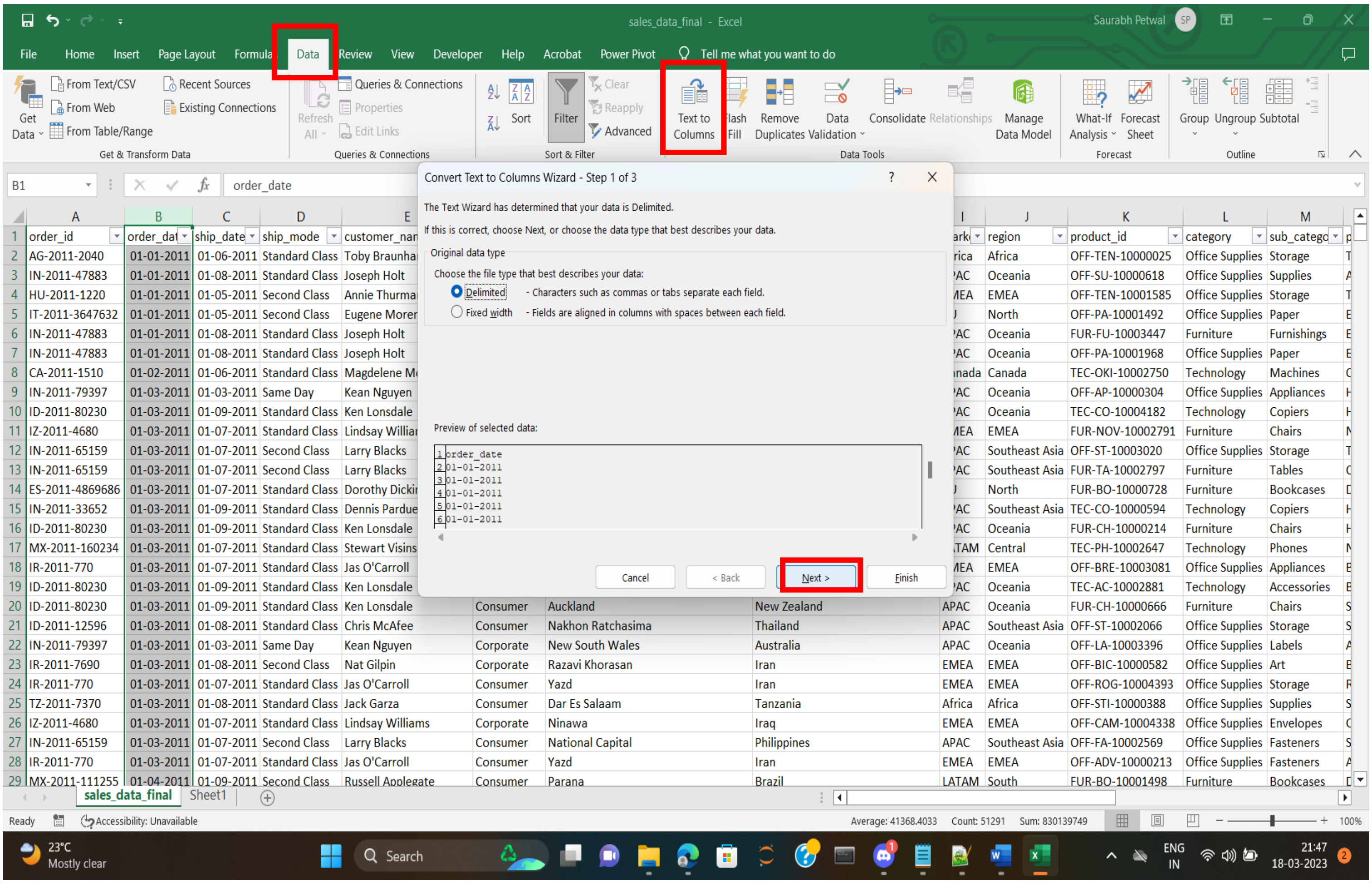
Documentation for Cleaning of Dataset Using Excel- Assignment No. 1

1. “order_date” and “ship_Date” columns in dataset are not in proper format. Some dates are in general format while rest are in date format. So changed the “order_date” and “ship_date” column format from “General” format to “date” format via the procedure shown below:

1.a) Select the **order_date** column (whole column)

1.b) Go to **Data** tab and then select **Text to columns**

1.c) Select • Delimited (by default) and click on **Next** button



Convert Text to Columns Wizard - Step 1 of 3

The Text Wizard has determined that your data is Delimited.

If this is correct, choose Next, or choose the data type that best describes your data.

Original data type

Choose the file type that best describes your data:

● Delimited - Characters such as commas or tabs separate each field.

○ Fixed width - Fields are aligned in columns with spaces between each field.

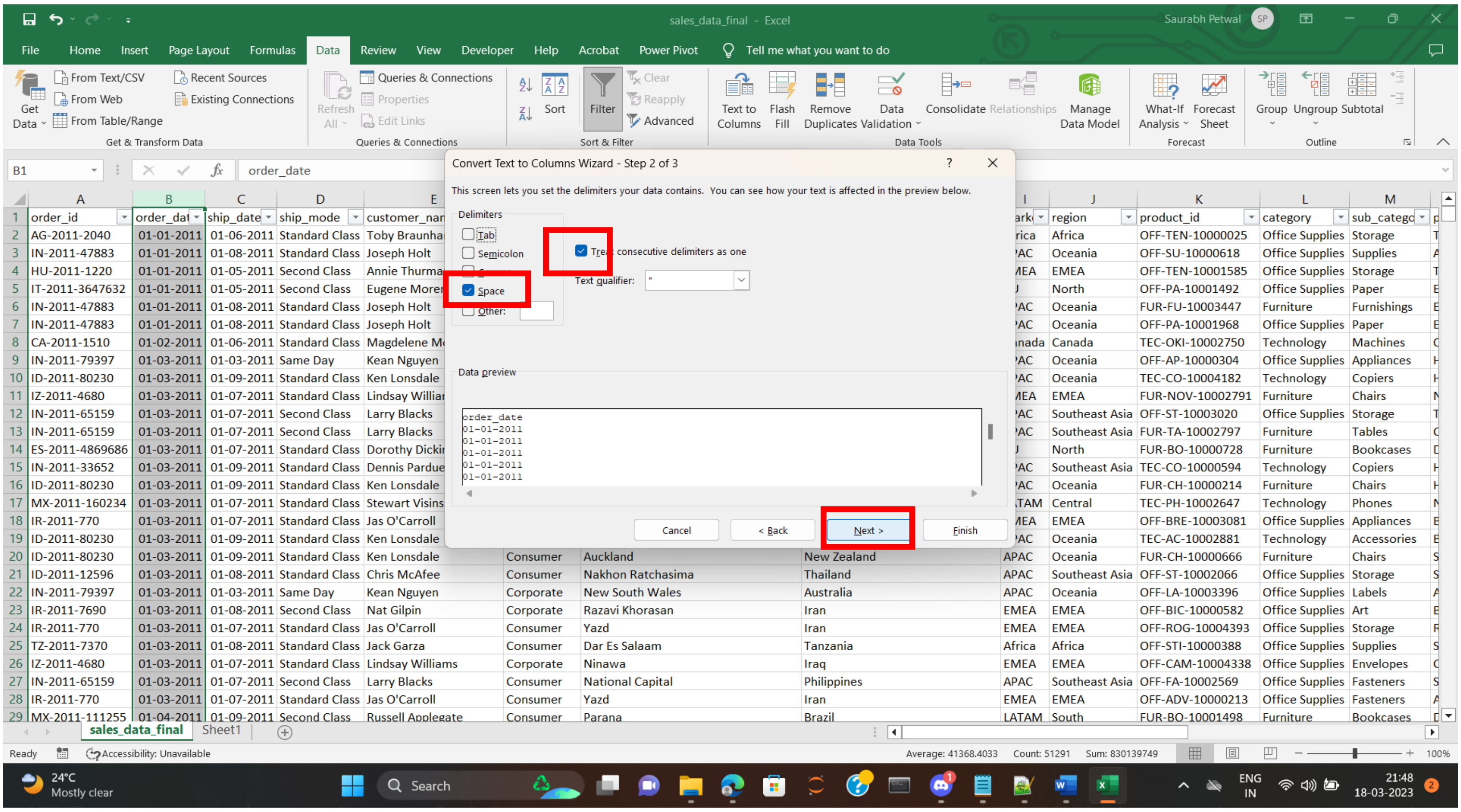
Preview of selected data:

```
1 order_date
2 01-01-2011
3 01-01-2011
4 01-01-2011
5 01-01-2011
6 01-01-2011
```

Cancel | < Back | Next > | Finish

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	order_id	order_dat	ship_date	ship_mode	customer_nan				ark	region	product_id	category	sub_catego
2	AG-2011-2040	01-01-2011	01-06-2011	Standard Class	Toby Braunha				rica	Africa	OFF-TEN-10000025	Office Supplies	Storage
3	IN-2011-47883	01-01-2011	01-08-2011	Standard Class	Joseph Holt				AC	Oceania	OFF-SU-10000618	Office Supplies	Supplies
4	HU-2011-1220	01-01-2011	01-05-2011	Second Class	Annie Thurma				MEA	EMEA	OFF-TEN-10001585	Office Supplies	Storage
5	IT-2011-3647632	01-01-2011	01-05-2011	Second Class	Eugene Morer				J	North	OFF-PA-10001492	Office Supplies	Paper
6	IN-2011-47883	01-01-2011	01-08-2011	Standard Class	Joseph Holt				AC	Oceania	FUR-FU-10003447	Furniture	Furnishings
7	IN-2011-47883	01-01-2011	01-08-2011	Standard Class	Joseph Holt				AC	Oceania	OFF-PA-10001968	Office Supplies	Paper
8	CA-2011-1510	01-02-2011	01-06-2011	Standard Class	Magdelene M				nada	Canada	TEC-OKI-10002750	Technology	Machines
9	IN-2011-79397	01-03-2011	01-03-2011	Same Day	Kean Nguyen				AC	Oceania	OFF-AP-10000304	Office Supplies	Appliances
10	ID-2011-80230	01-03-2011	01-09-2011	Standard Class	Ken Lonsdale				AC	Oceania	TEC-CO-10004182	Technology	Copiers
11	IZ-2011-4680	01-03-2011	01-07-2011	Standard Class	Lindsay Willia				MEA	EMEA	FUR-NOV-10002791	Furniture	Chairs
12	IN-2011-65159	01-03-2011	01-07-2011	Second Class	Larry Blacks				AC	Southeast Asia	OFF-ST-10003020	Office Supplies	Storage
13	IN-2011-65159	01-03-2011	01-07-2011	Second Class	Larry Blacks				AC	Southeast Asia	FUR-TA-10002797	Furniture	Tables
14	ES-2011-4869686	01-03-2011	01-07-2011	Standard Class	Dorothy Dicki				J	North	FUR-BO-10000728	Furniture	Bookcases
15	IN-2011-33652	01-03-2011	01-09-2011	Standard Class	Dennis Pardue				AC	Southeast Asia	TEC-CO-10000594	Technology	Copiers
16	ID-2011-80230	01-03-2011	01-09-2011	Standard Class	Ken Lonsdale				AC	Oceania	FUR-CH-10000214	Furniture	Chairs
17	MX-2011-160234	01-03-2011	01-07-2011	Standard Class	Stewart Visins				TAM	Central	TEC-PH-10002647	Technology	Phones
18	IR-2011-770	01-03-2011	01-07-2011	Standard Class	Jas O'Carroll				MEA	EMEA	OFF-BRE-10003081	Office Supplies	Appliances
19	ID-2011-80230	01-03-2011	01-09-2011	Standard Class	Ken Lonsdale				AC	Oceania	TEC-AC-10002881	Technology	Accessories
20	ID-2011-80230	01-03-2011	01-09-2011	Standard Class	Ken Lonsdale	Consumer	Auckland	New Zealand	APAC	Oceania	FUR-CH-10000666	Furniture	Chairs
21	ID-2011-12596	01-03-2011	01-08-2011	Standard Class	Chris McAfee	Consumer	Nakhon Ratchasima	Thailand	APAC	Southeast Asia	OFF-ST-10002066	Office Supplies	Storage
22	IN-2011-79397	01-03-2011	01-03-2011	Same Day	Kean Nguyen	Corporate	New South Wales	Australia	APAC	Oceania	OFF-LA-10003396	Office Supplies	Labels
23	IR-2011-7690	01-03-2011	01-08-2011	Second Class	Nat Gilpin	Corporate	Razavi Khorasan	Iran	EMEA	EMEA	OFF-BIC-10000582	Office Supplies	Art
24	IR-2011-770	01-03-2011	01-07-2011	Standard Class	Jas O'Carroll	Consumer	Yazd	Iran	EMEA	EMEA	OFF-ROG-10004393	Office Supplies	Storage
25	TZ-2011-7370	01-03-2011	01-08-2011	Standard Class	Jack Garza	Consumer	Dar Es Salaam	Tanzania	Africa	Africa	OFF-STI-10000388	Office Supplies	Supplies
26	IZ-2011-4680	01-03-2011	01-07-2011	Standard Class	Lindsay Williams	Corporate	Ninawa	Iraq	EMEA	EMEA	OFF-CAM-10004338	Office Supplies	Envelopes
27	IN-2011-65159	01-03-2011	01-07-2011	Second Class	Larry Blacks	Consumer	National Capital	Philippines	APAC	Southeast Asia	OFF-FA-10002569	Office Supplies	Fasteners
28	IR-2011-770	01-03-2011	01-07-2011	Standard Class	Jas O'Carroll	Consumer	Yazd	Iran	EMEA	EMEA	OFF-ADV-10000213	Office Supplies	Fasteners
29	MX-2011-111255	01-04-2011	01-09-2011	Second Class	Russell Applegate	Consumer	Parana	Brazil	LATAM	South	FUR-BO-10001498	Furniture	Bookcases

1.d) Check on √ **Space** and click on **Next** button



Convert Text to Columns Wizard - Step 2 of 3

This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.

Delimiters

- □ Tab
- □ Semicolon
- □ Comma
- ☑ Space
- □ Other:

☑ Treat consecutive delimiters as one

Text qualifier: "

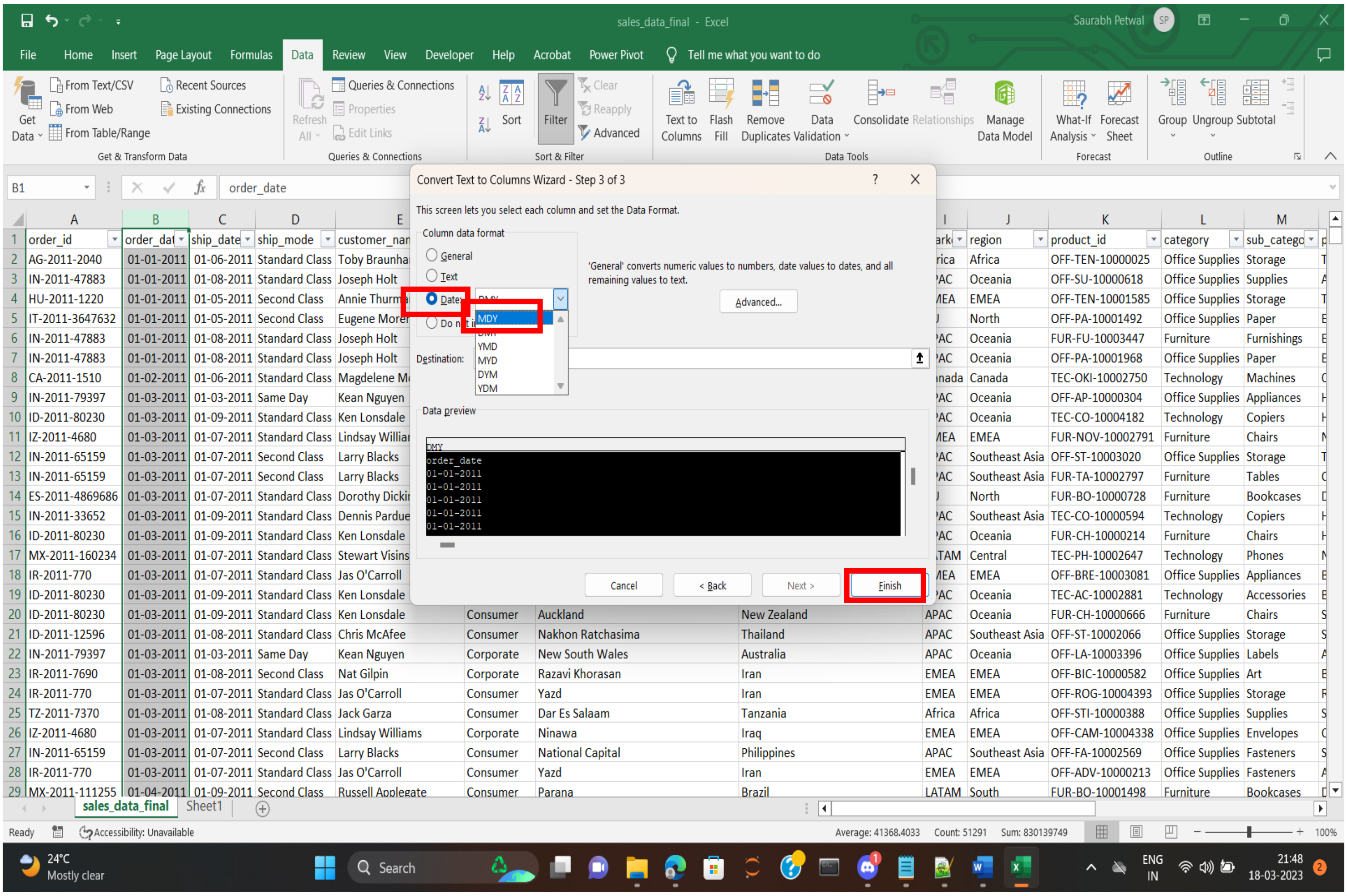
Data preview

```
order_date
01-01-2011
01-01-2011
01-01-2011
01-01-2011
01-01-2011
```

Cancel | < Back | Next > | Finish

	A	B	C	D	E				I	J	K	L	M	
1	order_id	order_dat	ship_date	ship_mode	customer_nan				ark	region	product_id	category	sub_catego	p
2	AG-2011-2040	01-01-2011	01-06-2011	Standard Class	Toby Braunha				rica	Africa	OFF-TEN-10000025	Office Supplies	Storage	T
3	IN-2011-47883	01-01-2011	01-08-2011	Standard Class	Joseph Holt				AC	Oceania	OFF-SU-10000618	Office Supplies	Supplies	A
4	HU-2011-1220	01-01-2011	01-05-2011	Second Class	Annie Thurma				IEA	EMEA	OFF-TEN-10001585	Office Supplies	Storage	T
5	IT-2011-3647632	01-01-2011	01-05-2011	Second Class	Eugene More				J	North	OFF-PA-10001492	Office Supplies	Paper	E
6	IN-2011-47883	01-01-2011	01-08-2011	Standard Class	Joseph Holt				AC	Oceania	FUR-FU-10003447	Furniture	Furnishings	E
7	IN-2011-47883	01-01-2011	01-08-2011	Standard Class	Joseph Holt				AC	Oceania	OFF-PA-10001968	Office Supplies	Paper	E
8	CA-2011-1510	01-02-2011	01-06-2011	Standard Class	Magdelene M				nada	Canada	TEC-OKI-10002750	Technology	Machines	C
9	IN-2011-79397	01-03-2011	01-03-2011	Same Day	Kean Nguyen				AC	Oceania	OFF-AP-10000304	Office Supplies	Appliances	H
10	ID-2011-80230	01-03-2011	01-09-2011	Standard Class	Ken Lonsdale				AC	Oceania	TEC-CO-10004182	Technology	Copiers	H
11	IZ-2011-4680	01-03-2011	01-07-2011	Standard Class	Lindsay Willian				IEA	EMEA	FUR-NOV-10002791	Furniture	Chairs	N
12	IN-2011-65159	01-03-2011	01-07-2011	Second Class	Larry Blacks				AC	Southeast Asia	OFF-ST-10003020	Office Supplies	Storage	T
13	IN-2011-65159	01-03-2011	01-07-2011	Second Class	Larry Blacks				AC	Southeast Asia	FUR-TA-10002797	Furniture	Tables	C
14	ES-2011-4869686	01-03-2011	01-07-2011	Standard Class	Dorothy Dickir				J	North	FUR-BO-10000728	Furniture	Bookcases	D
15	IN-2011-33652	01-03-2011	01-09-2011	Standard Class	Dennis Pardue				AC	Southeast Asia	TEC-CO-10000594	Technology	Copiers	H
16	ID-2011-80230	01-03-2011	01-09-2011	Standard Class	Ken Lonsdale				AC	Oceania	FUR-CH-10000214	Furniture	Chairs	H
17	MX-2011-160234	01-03-2011	01-07-2011	Standard Class	Stewart Visins				TAM	Central	TEC-PH-10002647	Technology	Phones	N
18	IR-2011-770	01-03-2011	01-07-2011	Standard Class	Jas O'Carroll				IEA	EMEA	OFF-BRE-10003081	Office Supplies	Appliances	E
19	ID-2011-80230	01-03-2011	01-09-2011	Standard Class	Ken Lonsdale				AC	Oceania	TEC-AC-10002881	Technology	Accessories	E
20	ID-2011-80230	01-03-2011	01-09-2011	Standard Class	Ken Lonsdale	Consumer	Auckland	New Zealand	APAC	Oceania	FUR-CH-10000666	Furniture	Chairs	S
21	ID-2011-12596	01-03-2011	01-08-2011	Standard Class	Chris McAfee	Consumer	Nakhon Ratchasima	Thailand	APAC	Southeast Asia	OFF-ST-10002066	Office Supplies	Storage	S
22	IN-2011-79397	01-03-2011	01-03-2011	Same Day	Kean Nguyen	Corporate	New South Wales	Australia	APAC	Oceania	OFF-LA-10003396	Office Supplies	Labels	A
23	IR-2011-7690	01-03-2011	01-08-2011	Second Class	Nat Gilpin	Corporate	Razavi Khorasan	Iran	EMEA	EMEA	OFF-BIC-10000582	Office Supplies	Art	E
24	IR-2011-770	01-03-2011	01-07-2011	Standard Class	Jas O'Carroll	Consumer	Yazd	Iran	EMEA	EMEA	OFF-ROG-10004393	Office Supplies	Storage	R
25	TZ-2011-7370	01-03-2011	01-08-2011	Standard Class	Jack Garza	Consumer	Dar Es Salaam	Tanzania	Africa	Africa	OFF-STI-10000388	Office Supplies	Supplies	S
26	IZ-2011-4680	01-03-2011	01-07-2011	Standard Class	Lindsay Williams	Corporate	Ninawa	Iraq	EMEA	EMEA	OFF-CAM-10004338	Office Supplies	Envelopes	C
27	IN-2011-65159	01-03-2011	01-07-2011	Second Class	Larry Blacks	Consumer	National Capital	Philippines	APAC	Southeast Asia	OFF-FA-10002569	Office Supplies	Fasteners	S
28	IR-2011-770	01-03-2011	01-07-2011	Standard Class	Jas O'Carroll	Consumer	Yazd	Iran	EMEA	EMEA	OFF-ADV-10000213	Office Supplies	Fasteners	A
29	MX-2011-111255	01-04-2011	01-09-2011	Second Class	Russell Applegate	Consumer	Parana	Brazil	LATAM	South	FUR-BO-10001498	Furniture	Bookcases	D

1.e) Check on √ **Date** and select **DMY** format from drop down and then click on Finish button



B1 | order_date

Convert Text to Columns Wizard - Step 3 of 3

This screen lets you select each column and set the Data Format.

Column data format

○ General

○ Text

● Date: DMY

○ Do not import column (skip)

MDY
DMY
YMD
MYD
DYM
YDM

'General' converts numeric values to numbers, date values to dates, and all remaining values to text.

Advanced...

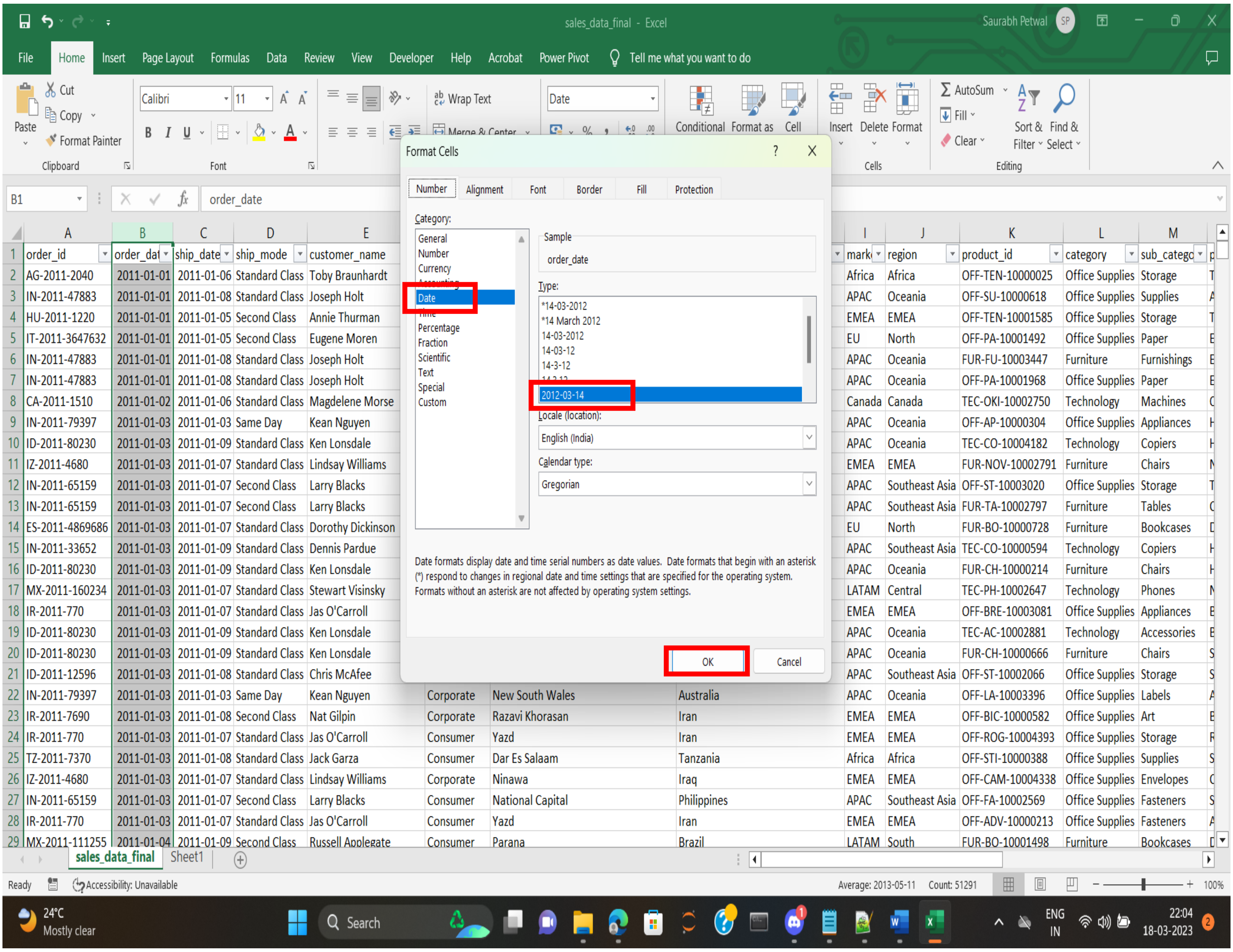
Destination:

Data preview

```
DMY
order_date
01-01-2011
01-01-2011
01-01-2011
01-01-2011
01-01-2011
```

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	order_id	order_dat	ship_date	ship_mode	customer_nan				ark	region	product_id	category	sub_catego
2	AG-2011-2040	01-01-2011	01-06-2011	Standard Class	Toby Braunha				rica	Africa	OFF-TEN-10000025	Office Supplies	Storage
3	IN-2011-47883	01-01-2011	01-08-2011	Standard Class	Joseph Holt				AC	Oceania	OFF-SU-10000618	Office Supplies	Supplies
4	HU-2011-1220	01-01-2011	01-05-2011	Second Class	Annie Thurma				MEA	EMEA	OFF-TEN-10001585	Office Supplies	Storage
5	IT-2011-3647632	01-01-2011	01-05-2011	Second Class	Eugene Morer				J	North	OFF-PA-10001492	Office Supplies	Paper
6	IN-2011-47883	01-01-2011	01-08-2011	Standard Class	Joseph Holt				AC	Oceania	FUR-FU-10003447	Furniture	Furnishings
7	IN-2011-47883	01-01-2011	01-08-2011	Standard Class	Joseph Holt				AC	Oceania	OFF-PA-10001968	Office Supplies	Paper
8	CA-2011-1510	01-02-2011	01-06-2011	Standard Class	Magdelene Mo				nada	Canada	TEC-OKI-10002750	Technology	Machines
9	IN-2011-79397	01-03-2011	01-03-2011	Same Day	Kean Nguyen				AC	Oceania	OFF-AP-10000304	Office Supplies	Appliances
10	ID-2011-80230	01-03-2011	01-09-2011	Standard Class	Ken Lonsdale				AC	Oceania	TEC-CO-10004182	Technology	Copiers
11	IZ-2011-4680	01-03-2011	01-07-2011	Standard Class	Lindsay Willian				MEA	EMEA	FUR-NOV-10002791	Furniture	Chairs
12	IN-2011-65159	01-03-2011	01-07-2011	Second Class	Larry Blacks				AC	Southeast Asia	OFF-ST-10003020	Office Supplies	Storage
13	IN-2011-65159	01-03-2011	01-07-2011	Second Class	Larry Blacks				AC	Southeast Asia	FUR-TA-10002797	Furniture	Tables
14	ES-2011-4869686	01-03-2011	01-07-2011	Standard Class	Dorothy Dickir				J	North	FUR-BO-10000728	Furniture	Bookcases
15	IN-2011-33652	01-03-2011	01-09-2011	Standard Class	Dennis Pardue				AC	Southeast Asia	TEC-CO-10000594	Technology	Copiers
16	ID-2011-80230	01-03-2011	01-09-2011	Standard Class	Ken Lonsdale				AC	Oceania	FUR-CH-10000214	Furniture	Chairs
17	MX-2011-160234	01-03-2011	01-07-2011	Standard Class	Stewart Visins				TAM	Central	TEC-PH-10002647	Technology	Phones
18	IR-2011-770	01-03-2011	01-07-2011	Standard Class	Jas O'Carroll				MEA	EMEA	OFF-BRE-10003081	Office Supplies	Appliances
19	ID-2011-80230	01-03-2011	01-09-2011	Standard Class	Ken Lonsdale				AC	Oceania	TEC-AC-10002881	Technology	Accessories
20	ID-2011-80230	01-03-2011	01-09-2011	Standard Class	Ken Lonsdale	Consumer	Auckland	New Zealand	APAC	Oceania	FUR-CH-10000666	Furniture	Chairs
21	ID-2011-12596	01-03-2011	01-08-2011	Standard Class	Chris McAfee	Consumer	Nakhon Ratchasima	Thailand	APAC	Southeast Asia	OFF-ST-10002066	Office Supplies	Storage
22	IN-2011-79397	01-03-2011	01-03-2011	Same Day	Kean Nguyen	Corporate	New South Wales	Australia	APAC	Oceania	OFF-LA-10003396	Office Supplies	Labels
23	IR-2011-7690	01-03-2011	01-08-2011	Second Class	Nat Gilpin	Corporate	Razavi Khorasan	Iran	EMEA	EMEA	OFF-BIC-10000582	Office Supplies	Art
24	IR-2011-770	01-03-2011	01-07-2011	Standard Class	Jas O'Carroll	Consumer	Yazd	Iran	EMEA	EMEA	OFF-ROG-10004393	Office Supplies	Storage
25	TZ-2011-7370	01-03-2011	01-08-2011	Standard Class	Jack Garza	Consumer	Dar Es Salaam	Tanzania	Africa	Africa	OFF-STI-10000388	Office Supplies	Supplies
26	IZ-2011-4680	01-03-2011	01-07-2011	Standard Class	Lindsay Williams	Corporate	Ninawa	Iraq	EMEA	EMEA	OFF-CAM-10004338	Office Supplies	Envelopes
27	IN-2011-65159	01-03-2011	01-07-2011	Second Class	Larry Blacks	Consumer	National Capital	Philippines	APAC	Southeast Asia	OFF-FA-10002569	Office Supplies	Fasteners
28	IR-2011-770	01-03-2011	01-07-2011	Standard Class	Jas O'Carroll	Consumer	Yazd	Iran	EMEA	EMEA	OFF-ADV-10000213	Office Supplies	Fasteners
29	MX-2011-111255	01-04-2011	01-09-2011	Second Class	Russell Applegate	Consumer	Parana	Brazil	LATAM	South	FUR-BO-10001498	Furniture	Bookcases

1.f) Now, Go to Format type (**ctrl +1** : shortcut key) and select date format as YYYY-MM-DD, as snow flake takes date in the same format



Format Cells

Number | Alignment | Font | Border | Fill | Protection

Category:
General
Number
Currency
Accounting
Date
Time
Percentage
Fraction
Scientific
Text
Special
Custom

Sample
order_date

Type:
*14-03-2012
*14 March 2012
14-03-2012
14-03-12
14-3-12
14-3-12
2012-03-14

Locale (location): English (India)

Calendar type: Gregorian

Date formats display date and time serial numbers as date values. Date formats that begin with an asterisk (*) respond to changes in regional date and time settings that are specified for the operating system. Formats without an asterisk are not affected by operating system settings.

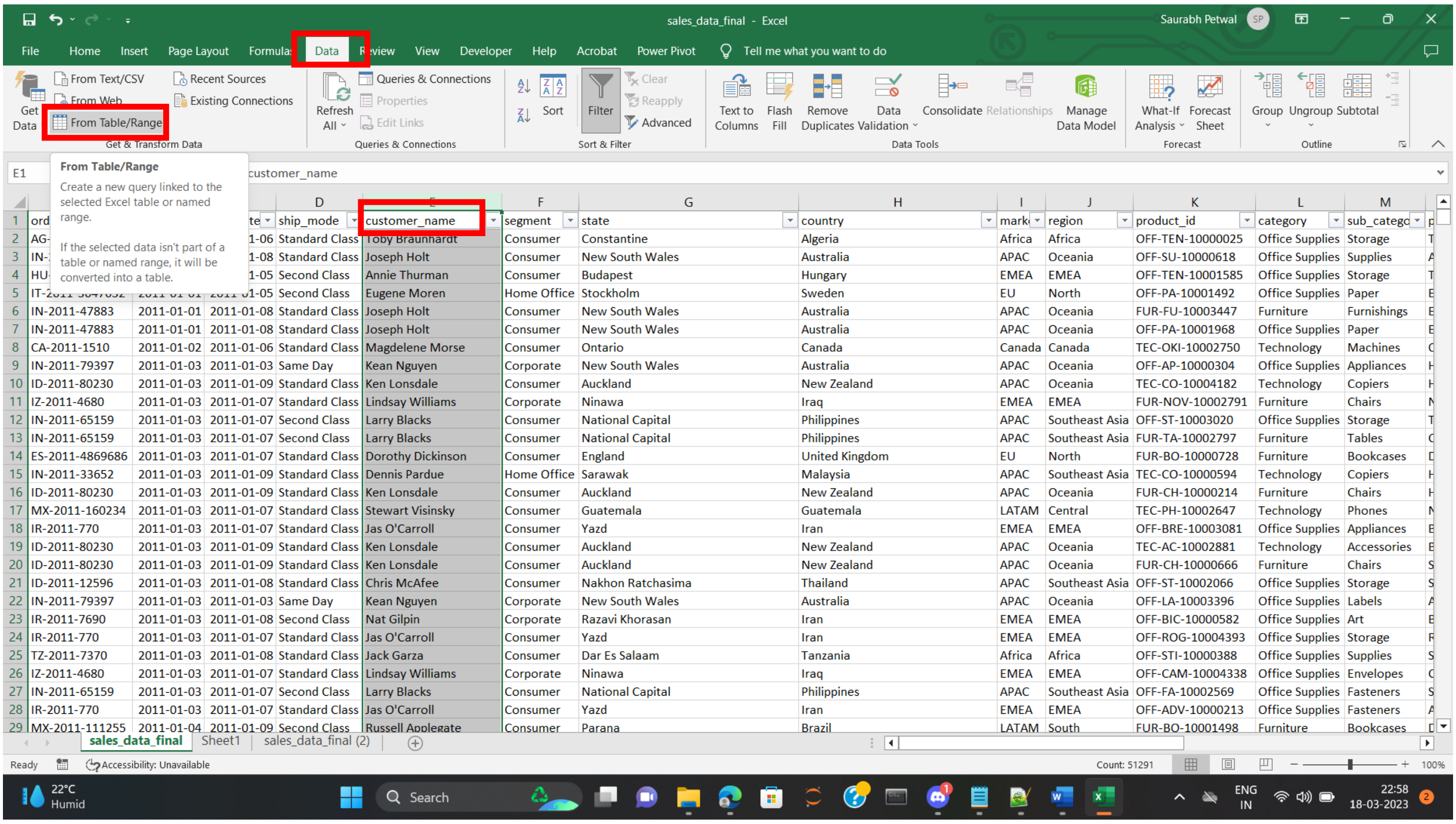
OK | Cancel

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	order_id	order_dat	ship_date	ship_mode	customer_name				mark	region	product_id	category	sub_catego
2	AG-2011-2040	2011-01-01	2011-01-06	Standard Class	Toby Braunhardt				Africa	Africa	OFF-TEN-10000025	Office Supplies	Storage
3	IN-2011-47883	2011-01-01	2011-01-08	Standard Class	Joseph Holt				APAC	Oceania	OFF-SU-10000618	Office Supplies	Supplies
4	HU-2011-1220	2011-01-01	2011-01-05	Second Class	Annie Thurman				EMEA	EMEA	OFF-TEN-10001585	Office Supplies	Storage
5	IT-2011-3647632	2011-01-01	2011-01-05	Second Class	Eugene Moren				EU	North	OFF-PA-10001492	Office Supplies	Paper
6	IN-2011-47883	2011-01-01	2011-01-08	Standard Class	Joseph Holt				APAC	Oceania	FUR-FU-10003447	Furniture	Furnishings
7	IN-2011-47883	2011-01-01	2011-01-08	Standard Class	Joseph Holt				APAC	Oceania	OFF-PA-10001968	Office Supplies	Paper
8	CA-2011-1510	2011-01-02	2011-01-06	Standard Class	Magdelene Morse				Canada	Canada	TEC-OKI-10002750	Technology	Machines
9	IN-2011-79397	2011-01-03	2011-01-03	Same Day	Kean Nguyen				APAC	Oceania	OFF-AP-10000304	Office Supplies	Appliances
10	ID-2011-80230	2011-01-03	2011-01-09	Standard Class	Ken Lonsdale				APAC	Oceania	TEC-CO-10004182	Technology	Copiers
11	IZ-2011-4680	2011-01-03	2011-01-07	Standard Class	Lindsay Williams				EMEA	EMEA	FUR-NOV-10002791	Furniture	Chairs
12	IN-2011-65159	2011-01-03	2011-01-07	Second Class	Larry Blacks				APAC	Southeast Asia	OFF-ST-10003020	Office Supplies	Storage
13	IN-2011-65159	2011-01-03	2011-01-07	Second Class	Larry Blacks				APAC	Southeast Asia	FUR-TA-10002797	Furniture	Tables
14	ES-2011-4869686	2011-01-03	2011-01-07	Standard Class	Dorothy Dickinson				EU	North	FUR-BO-10000728	Furniture	Bookcases
15	IN-2011-33652	2011-01-03	2011-01-09	Standard Class	Dennis Pardue				APAC	Southeast Asia	TEC-CO-10000594	Technology	Copiers
16	ID-2011-80230	2011-01-03	2011-01-09	Standard Class	Ken Lonsdale				APAC	Oceania	FUR-CH-10000214	Furniture	Chairs
17	MX-2011-160234	2011-01-03	2011-01-07	Standard Class	Stewart Visinsky				LATAM	Central	TEC-PH-10002647	Technology	Phones
18	IR-2011-770	2011-01-03	2011-01-07	Standard Class	Jas O'Carroll				EMEA	EMEA	OFF-BRE-10003081	Office Supplies	Appliances
19	ID-2011-80230	2011-01-03	2011-01-09	Standard Class	Ken Lonsdale				APAC	Oceania	TEC-AC-10002881	Technology	Accessories
20	ID-2011-80230	2011-01-03	2011-01-09	Standard Class	Ken Lonsdale				APAC	Oceania	FUR-CH-10000666	Furniture	Chairs
21	ID-2011-12596	2011-01-03	2011-01-08	Standard Class	Chris McAfee				APAC	Southeast Asia	OFF-ST-10002066	Office Supplies	Storage
22	IN-2011-79397	2011-01-03	2011-01-03	Same Day	Kean Nguyen	Corporate	New South Wales	Australia	APAC	Oceania	OFF-LA-10003396	Office Supplies	Labels
23	IR-2011-7690	2011-01-03	2011-01-08	Second Class	Nat Gilpin	Corporate	Razavi Khorasan	Iran	EMEA	EMEA	OFF-BIC-10000582	Office Supplies	Art
24	IR-2011-770	2011-01-03	2011-01-07	Standard Class	Jas O'Carroll	Consumer	Yazd	Iran	EMEA	EMEA	OFF-ROG-10004393	Office Supplies	Storage
25	TZ-2011-7370	2011-01-03	2011-01-08	Standard Class	Jack Garza	Consumer	Dar Es Salaam	Tanzania	Africa	Africa	OFF-STI-10000388	Office Supplies	Supplies
26	IZ-2011-4680	2011-01-03	2011-01-07	Standard Class	Lindsay Williams	Corporate	Ninawa	Iraq	EMEA	EMEA	OFF-CAM-10004338	Office Supplies	Envelopes
27	IN-2011-65159	2011-01-03	2011-01-07	Second Class	Larry Blacks	Consumer	National Capital	Philippines	APAC	Southeast Asia	OFF-FA-10002569	Office Supplies	Fasteners
28	IR-2011-770	2011-01-03	2011-01-07	Standard Class	Jas O'Carroll	Consumer	Yazd	Iran	EMEA	EMEA	OFF-ADV-10000213	Office Supplies	Fasteners
29	MX-2011-111255	2011-01-04	2011-01-09	Second Class	Russell Applegate	Consumer	Parana	Brazil	LATAM	South	FUR-BO-10001498	Furniture	Bookcases

Note: This is the effective way to turn the format to data type using excel. Similarly follow same steps for column “ship_date”.

2. Using Custom filter checking data in columns "customer_name", "product_name", "state" as these columns have special characters so removed the special characters and UTF 8 characters step wise step. Using below procedure we can clean the data in alpha numeric datatype but since it also contain UTF8 characters and some latin and german characters, so to keep in mind to not to loose data we will manually remove and replace them. We will create a Flag for rows having special characters and can remove them manually using replace function (**ctrl +f** >> then click Replace tab).

   2.a) Select the whole column **"customer_name"** and then click on **"Data"** tab. Now select From **Table/Range**



From Table/Range

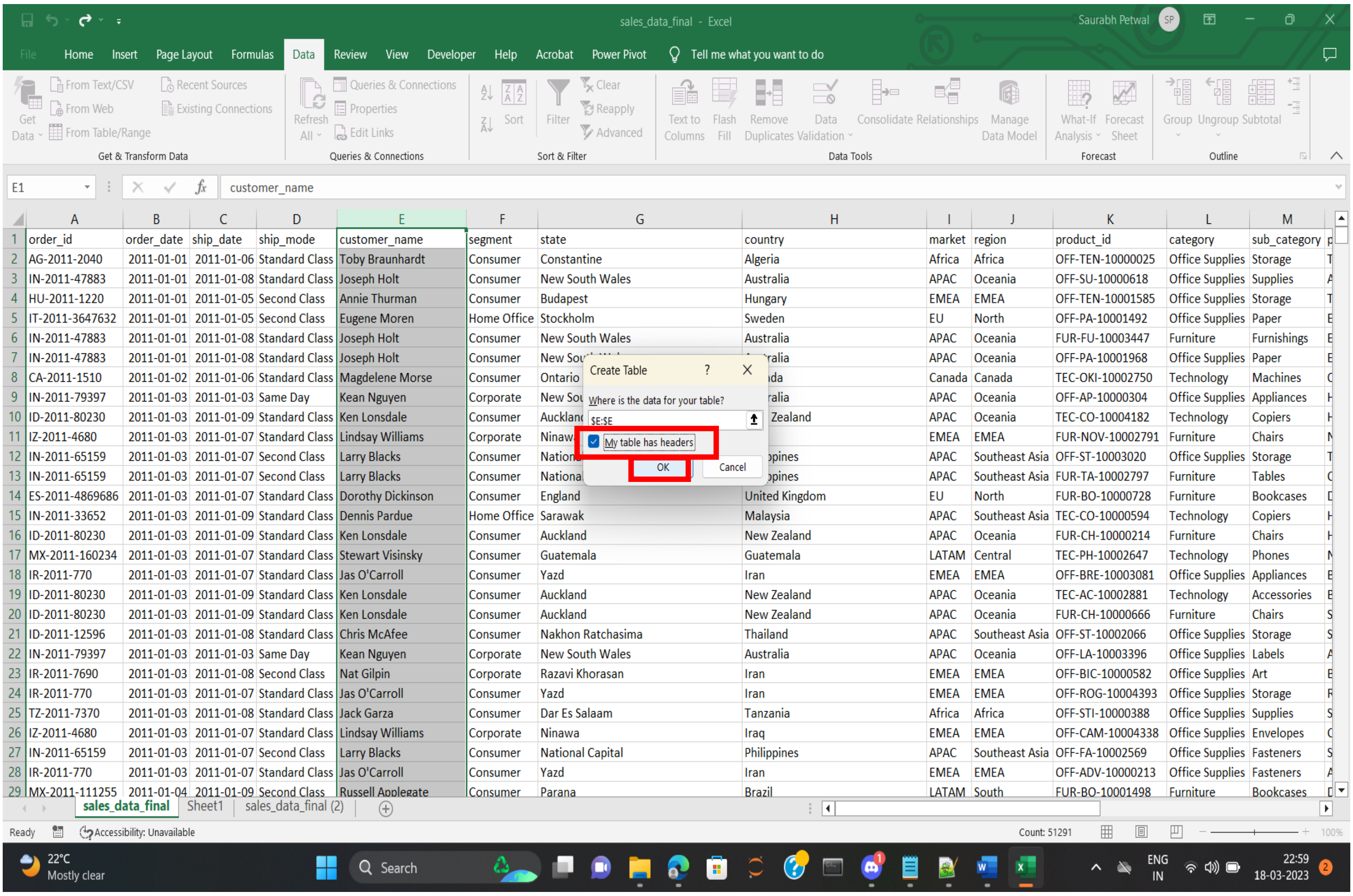
Create a new query linked to the selected Excel table or named range.

If the selected data isn't part of a table or named range, it will be converted into a table.

E1 customer_name

				D	E	F	G	H	I	J	K	L	M	
1	ord		te	ship_mode	customer_name	segment	state	country	mark	region	product_id	category	sub_catego	p
2	AG-		1-06	Standard Class	Toby Braunhardt	Consumer	Constantine	Algeria	Africa	Africa	OFF-TEN-10000025	Office Supplies	Storage	T
3	IN-		1-08	Standard Class	Joseph Holt	Consumer	New South Wales	Australia	APAC	Oceania	OFF-SU-10000618	Office Supplies	Supplies	A
4	HU-		1-05	Second Class	Annie Thurman	Consumer	Budapest	Hungary	EMEA	EMEA	OFF-TEN-10001585	Office Supplies	Storage	T
5	IT-2011-3047052	2011-01-01	2011-01-05	Second Class	Eugene Moren	Home Office	Stockholm	Sweden	EU	North	OFF-PA-10001492	Office Supplies	Paper	E
6	IN-2011-47883	2011-01-01	2011-01-08	Standard Class	Joseph Holt	Consumer	New South Wales	Australia	APAC	Oceania	FUR-FU-10003447	Furniture	Furnishings	E
7	IN-2011-47883	2011-01-01	2011-01-08	Standard Class	Joseph Holt	Consumer	New South Wales	Australia	APAC	Oceania	OFF-PA-10001968	Office Supplies	Paper	E
8	CA-2011-1510	2011-01-02	2011-01-06	Standard Class	Magdelene Morse	Consumer	Ontario	Canada	Canada	Canada	TEC-OKI-10002750	Technology	Machines	C
9	IN-2011-79397	2011-01-03	2011-01-03	Same Day	Kean Nguyen	Corporate	New South Wales	Australia	APAC	Oceania	OFF-AP-10000304	Office Supplies	Appliances	H
10	ID-2011-80230	2011-01-03	2011-01-09	Standard Class	Ken Lonsdale	Consumer	Auckland	New Zealand	APAC	Oceania	TEC-CO-10004182	Technology	Copiers	H
11	IZ-2011-4680	2011-01-03	2011-01-07	Standard Class	Lindsay Williams	Corporate	Ninawa	Iraq	EMEA	EMEA	FUR-NOV-10002791	Furniture	Chairs	N
12	IN-2011-65159	2011-01-03	2011-01-07	Second Class	Larry Blacks	Consumer	National Capital	Philippines	APAC	Southeast Asia	OFF-ST-10003020	Office Supplies	Storage	T
13	IN-2011-65159	2011-01-03	2011-01-07	Second Class	Larry Blacks	Consumer	National Capital	Philippines	APAC	Southeast Asia	FUR-TA-10002797	Furniture	Tables	C
14	ES-2011-4869686	2011-01-03	2011-01-07	Standard Class	Dorothy Dickinson	Consumer	England	United Kingdom	EU	North	FUR-BO-10000728	Furniture	Bookcases	D
15	IN-2011-33652	2011-01-03	2011-01-09	Standard Class	Dennis Pardue	Home Office	Sarawak	Malaysia	APAC	Southeast Asia	TEC-CO-10000594	Technology	Copiers	H
16	ID-2011-80230	2011-01-03	2011-01-09	Standard Class	Ken Lonsdale	Consumer	Auckland	New Zealand	APAC	Oceania	FUR-CH-10000214	Furniture	Chairs	H
17	MX-2011-160234	2011-01-03	2011-01-07	Standard Class	Stewart Visinsky	Consumer	Guatemala	Guatemala	LATAM	Central	TEC-PH-10002647	Technology	Phones	N
18	IR-2011-770	2011-01-03	2011-01-07	Standard Class	Jas O'Carroll	Consumer	Yazd	Iran	EMEA	EMEA	OFF-BRE-10003081	Office Supplies	Appliances	E
19	ID-2011-80230	2011-01-03	2011-01-09	Standard Class	Ken Lonsdale	Consumer	Auckland	New Zealand	APAC	Oceania	TEC-AC-10002881	Technology	Accessories	E
20	ID-2011-80230	2011-01-03	2011-01-09	Standard Class	Ken Lonsdale	Consumer	Auckland	New Zealand	APAC	Oceania	FUR-CH-10000666	Furniture	Chairs	S
21	ID-2011-12596	2011-01-03	2011-01-08	Standard Class	Chris McAfee	Consumer	Nakhon Ratchasima	Thailand	APAC	Southeast Asia	OFF-ST-10002066	Office Supplies	Storage	S
22	IN-2011-79397	2011-01-03	2011-01-03	Same Day	Kean Nguyen	Corporate	New South Wales	Australia	APAC	Oceania	OFF-LA-10003396	Office Supplies	Labels	A
23	IR-2011-7690	2011-01-03	2011-01-08	Second Class	Nat Gilpin	Corporate	Razavi Khorasan	Iran	EMEA	EMEA	OFF-BIC-10000582	Office Supplies	Art	E
24	IR-2011-770	2011-01-03	2011-01-07	Standard Class	Jas O'Carroll	Consumer	Yazd	Iran	EMEA	EMEA	OFF-ROG-10004393	Office Supplies	Storage	R
25	TZ-2011-7370	2011-01-03	2011-01-08	Standard Class	Jack Garza	Consumer	Dar Es Salaam	Tanzania	Africa	Africa	OFF-STI-10000388	Office Supplies	Supplies	S
26	IZ-2011-4680	2011-01-03	2011-01-07	Standard Class	Lindsay Williams	Corporate	Ninawa	Iraq	EMEA	EMEA	OFF-CAM-10004338	Office Supplies	Envelopes	C
27	IN-2011-65159	2011-01-03	2011-01-07	Second Class	Larry Blacks	Consumer	National Capital	Philippines	APAC	Southeast Asia	OFF-FA-10002569	Office Supplies	Fasteners	S
28	IR-2011-770	2011-01-03	2011-01-07	Standard Class	Jas O'Carroll	Consumer	Yazd	Iran	EMEA	EMEA	OFF-ADV-10000213	Office Supplies	Fasteners	A
29	MX-2011-111255	2011-01-04	2011-01-09	Second Class	Russell Applegate	Consumer	Parana	Brazil	LATAM	South	FUR-BO-10001498	Furniture	Bookcases	D

2.b) Check on popup **My table has headers** and then click on **OK**



E1 | customer_name

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	order_id	order_date	ship_date	ship_mode	customer_name	segment	state	country	market	region	product_id	category	sub_category
2	AG-2011-2040	2011-01-01	2011-01-06	Standard Class	Toby Braunhardt	Consumer	Constantine	Algeria	Africa	Africa	OFF-TEN-10000025	Office Supplies	Storage
3	IN-2011-47883	2011-01-01	2011-01-08	Standard Class	Joseph Holt	Consumer	New South Wales	Australia	APAC	Oceania	OFF-SU-10000618	Office Supplies	Supplies
4	HU-2011-1220	2011-01-01	2011-01-05	Second Class	Annie Thurman	Consumer	Budapest	Hungary	EMEA	EMEA	OFF-TEN-10001585	Office Supplies	Storage
5	IT-2011-3647632	2011-01-01	2011-01-05	Second Class	Eugene Moren	Home Office	Stockholm	Sweden	EU	North	OFF-PA-10001492	Office Supplies	Paper
6	IN-2011-47883	2011-01-01	2011-01-08	Standard Class	Joseph Holt	Consumer	New South Wales	Australia	APAC	Oceania	FUR-FU-10003447	Furniture	Furnishings
7	IN-2011-47883	2011-01-01	2011-01-08	Standard Class	Joseph Holt	Consumer	New Sou	ralia	APAC	Oceania	OFF-PA-10001968	Office Supplies	Paper
8	CA-2011-1510	2011-01-02	2011-01-06	Standard Class	Magdelene Morse	Consumer	Ontario	da	Canada	Canada	TEC-OKI-10002750	Technology	Machines
9	IN-2011-79397	2011-01-03	2011-01-03	Same Day	Kean Nguyen	Corporate	New Sou	ralia	APAC	Oceania	OFF-AP-10000304	Office Supplies	Appliances
10	ID-2011-80230	2011-01-03	2011-01-09	Standard Class	Ken Lonsdale	Consumer	Aucklanc	Zealand	APAC	Oceania	TEC-CO-10004182	Technology	Copiers
11	IZ-2011-4680	2011-01-03	2011-01-07	Standard Class	Lindsay Williams	Corporate	Ninaw		EMEA	EMEA	FUR-NOV-10002791	Furniture	Chairs
12	IN-2011-65159	2011-01-03	2011-01-07	Second Class	Larry Blacks	Consumer	Nation	ppines	APAC	Southeast Asia	OFF-ST-10003020	Office Supplies	Storage
13	IN-2011-65159	2011-01-03	2011-01-07	Second Class	Larry Blacks	Consumer	National	ppines	APAC	Southeast Asia	FUR-TA-10002797	Furniture	Tables
14	ES-2011-4869686	2011-01-03	2011-01-07	Standard Class	Dorothy Dickinson	Consumer	England	United Kingdom	EU	North	FUR-BO-10000728	Furniture	Bookcases
15	IN-2011-33652	2011-01-03	2011-01-09	Standard Class	Dennis Pardue	Home Office	Sarawak	Malaysia	APAC	Southeast Asia	TEC-CO-10000594	Technology	Copiers
16	ID-2011-80230	2011-01-03	2011-01-09	Standard Class	Ken Lonsdale	Consumer	Auckland	New Zealand	APAC	Oceania	FUR-CH-10000214	Furniture	Chairs
17	MX-2011-160234	2011-01-03	2011-01-07	Standard Class	Stewart Visinsky	Consumer	Guatemala	Guatemala	LATAM	Central	TEC-PH-10002647	Technology	Phones
18	IR-2011-770	2011-01-03	2011-01-07	Standard Class	Jas O'Carroll	Consumer	Yazd	Iran	EMEA	EMEA	OFF-BRE-10003081	Office Supplies	Appliances
19	ID-2011-80230	2011-01-03	2011-01-09	Standard Class	Ken Lonsdale	Consumer	Auckland	New Zealand	APAC	Oceania	TEC-AC-10002881	Technology	Accessories
20	ID-2011-80230	2011-01-03	2011-01-09	Standard Class	Ken Lonsdale	Consumer	Auckland	New Zealand	APAC	Oceania	FUR-CH-10000666	Furniture	Chairs
21	ID-2011-12596	2011-01-03	2011-01-08	Standard Class	Chris McAfee	Consumer	Nakhon Ratchasima	Thailand	APAC	Southeast Asia	OFF-ST-10002066	Office Supplies	Storage
22	IN-2011-79397	2011-01-03	2011-01-03	Same Day	Kean Nguyen	Corporate	New South Wales	Australia	APAC	Oceania	OFF-LA-10003396	Office Supplies	Labels
23	IR-2011-7690	2011-01-03	2011-01-08	Second Class	Nat Gilpin	Corporate	Razavi Khorasan	Iran	EMEA	EMEA	OFF-BIC-10000582	Office Supplies	Art
24	IR-2011-770	2011-01-03	2011-01-07	Standard Class	Jas O'Carroll	Consumer	Yazd	Iran	EMEA	EMEA	OFF-ROG-10004393	Office Supplies	Storage
25	TZ-2011-7370	2011-01-03	2011-01-08	Standard Class	Jack Garza	Consumer	Dar Es Salaam	Tanzania	Africa	Africa	OFF-STI-10000388	Office Supplies	Supplies
26	IZ-2011-4680	2011-01-03	2011-01-07	Standard Class	Lindsay Williams	Corporate	Ninawa	Iraq	EMEA	EMEA	OFF-CAM-10004338	Office Supplies	Envelopes
27	IN-2011-65159	2011-01-03	2011-01-07	Second Class	Larry Blacks	Consumer	National Capital	Philippines	APAC	Southeast Asia	OFF-FA-10002569	Office Supplies	Fasteners
28	IR-2011-770	2011-01-03	2011-01-07	Standard Class	Jas O'Carroll	Consumer	Yazd	Iran	EMEA	EMEA	OFF-ADV-10000213	Office Supplies	Fasteners
29	MX-2011-111255	2011-01-04	2011-01-09	Second Class	Russell Applegate	Consumer	Parana	Brazil	LATAM	South	FUR-BO-10001498	Furniture	Bookcases

Create Table ? ×

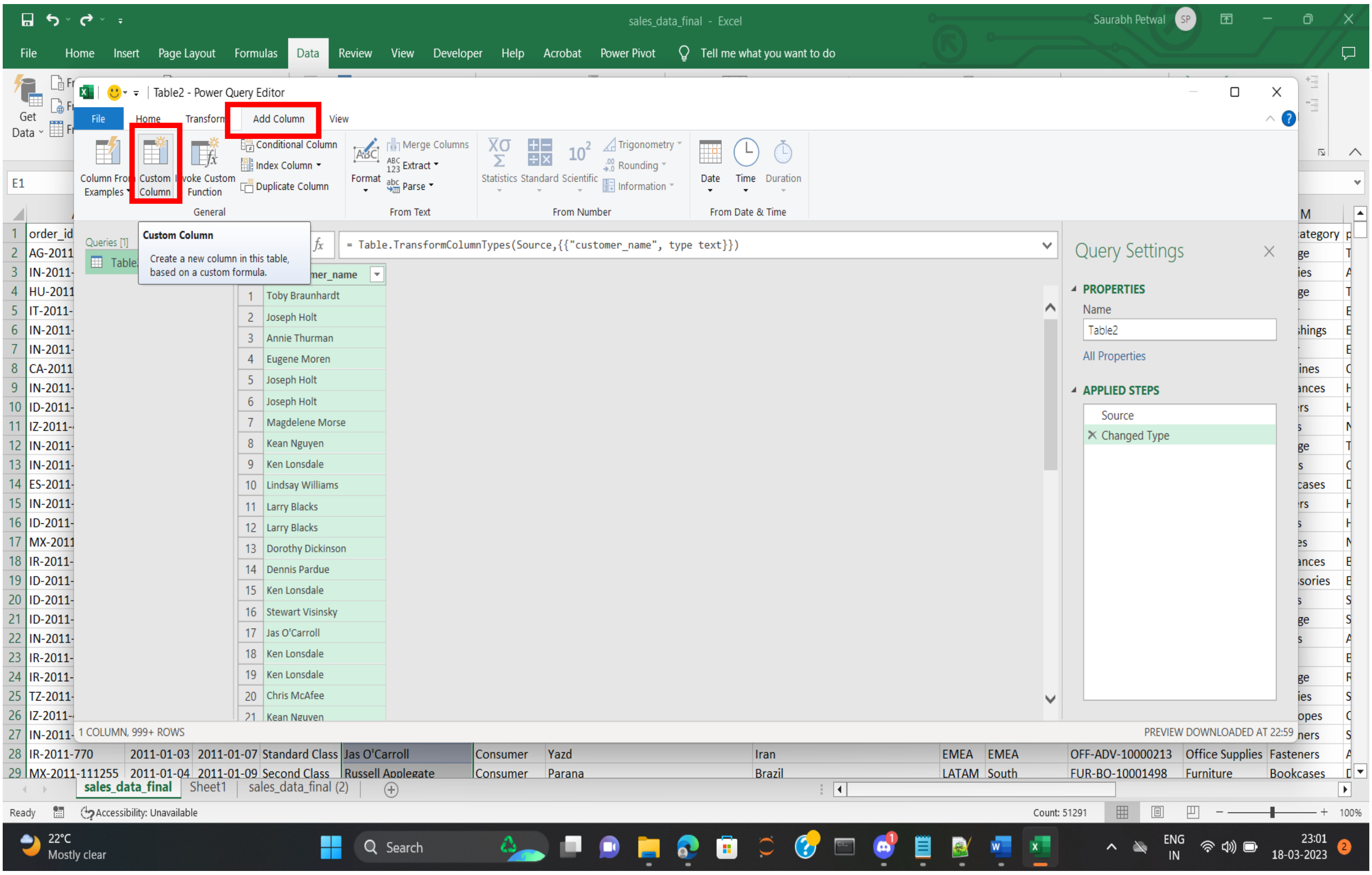
Where is the data for your table?

$E:$E

☑ My table has headers

OK Cancel

2.c) Click on **“Add column”** tab then click on **“Custom Column”**



Custom Column

Create a new column in this table, based on a custom formula.

```
= Table.TransformColumnTypes(Source,{{"customer_name", type text}})
```

	customer_name
1	Toby Braunhardt
2	Joseph Holt
3	Annie Thurman
4	Eugene Moren
5	Joseph Holt
6	Joseph Holt
7	Magdelene Morse
8	Kean Nguyen
9	Ken Lonsdale
10	Lindsay Williams
11	Larry Blacks
12	Larry Blacks
13	Dorothy Dickinson
14	Dennis Pardue
15	Ken Lonsdale
16	Stewart Visinsky
17	Jas O'Carroll
18	Ken Lonsdale
19	Ken Lonsdale
20	Chris McAfee
21	Kean Nguyen

Name: Table2

APPLIED STEPS: Source; Changed Type

1 COLUMN, 999+ ROWS

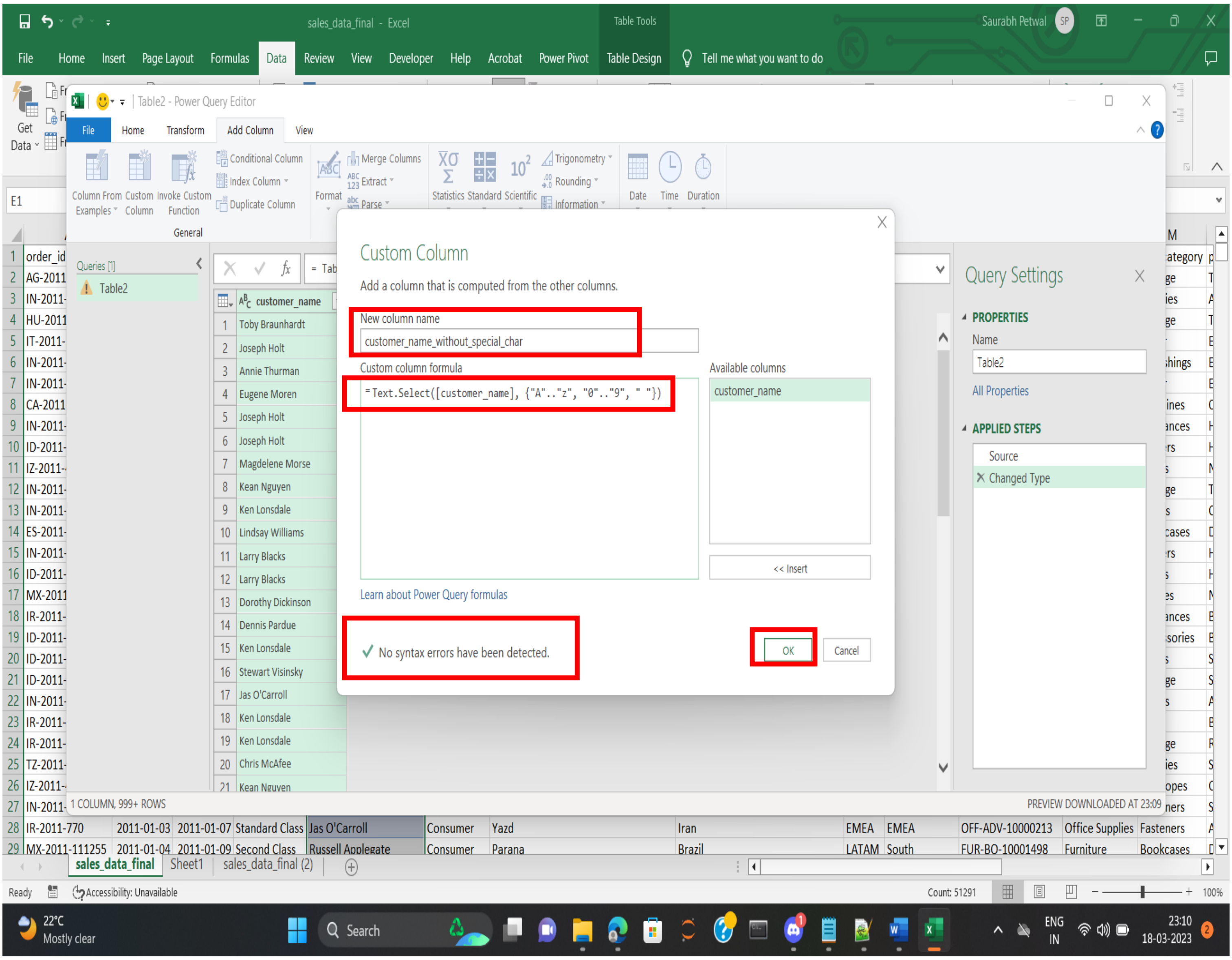
PREVIEW DOWNLOADED AT 22:59

2.d) Now type a new column name, then write a syntax to clean the existing column and the values will be saved to new_column, syntax used to convert the data to alphanumeric datatype is as:

= Text.Select([customer_name], {“A”..”z”, “0”..”9”, “ “})

Here,

- “customer_name” is the existing column in the table to which we are cleaning and creating a new column to save values in it, it is named below as “customer_name_with_special_char”
- “A”..”z” in the syntax is used for all values of “A-z and a-z “ since ASCII char are in sequence A,B,C…..Y,Z,a,b…y,z (capital then small case) so it gives range for capital as well as small case alphabets
- “0”..”9” in the syntax used for all numeric values from 0 to 9
- “ “ (white space) is used for white space between two words as we have data with space separated as first and last name of the customer in customer_name column. If we don’t use this “ “ in our sytntax then forst and last name will not having space between them in the new column generated by cleaning of data i.e “customer_name_with_special_char”



Custom Column

Add a column that is computed from the other columns.

New column name

customer_name_without_special_char

Custom column formula

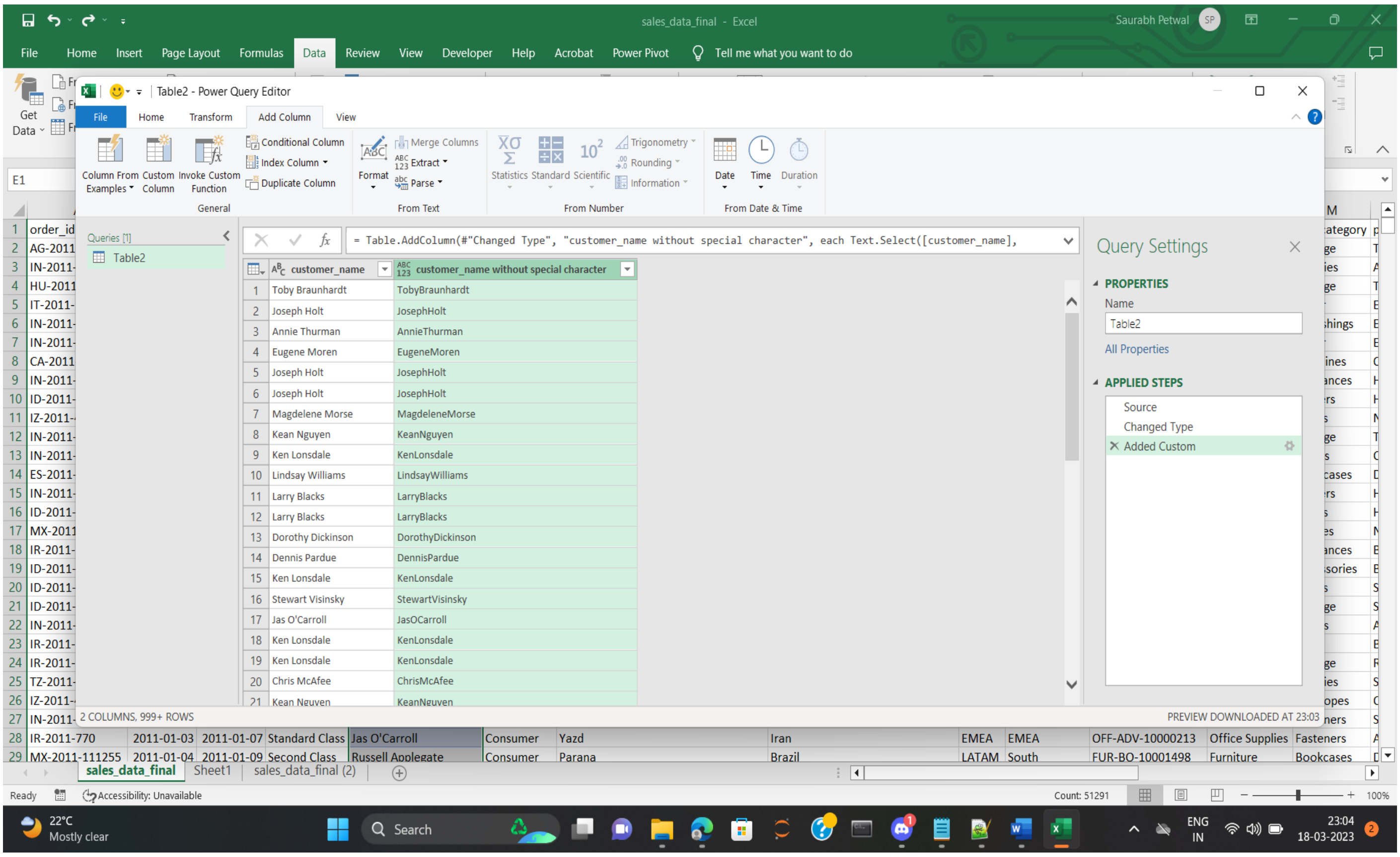
```
= Text.Select([customer_name], {"A".."z", "0".."9", " "})
```

Available columns

customer_name

✓ No syntax errors have been detected.

2.e) New column will be generated from existing column with cleaned data and can be used further.



```
= Table.AddColumn(#"Changed Type", "customer_name without special character", each Text.Select([customer_name],
```

	customer_name	customer_name without special character
1	Toby Braunhardt	TobyBraunhardt
2	Joseph Holt	JosephHolt
3	Annie Thurman	AnnieThurman
4	Eugene Moren	EugeneMoren
5	Joseph Holt	JosephHolt
6	Joseph Holt	JosephHolt
7	Magdelene Morse	MagdeleneMorse
8	Kean Nguyen	KeanNguyen
9	Ken Lonsdale	KenLonsdale
10	Lindsay Williams	LindsayWilliams
11	Larry Blacks	LarryBlacks
12	Larry Blacks	LarryBlacks
13	Dorothy Dickinson	DorothyDickinson
14	Dennis Pardue	DennisPardue
15	Ken Lonsdale	KenLonsdale
16	Stewart Visinsky	StewartVisinsky
17	Jas O'Carroll	JasOCarroll
18	Ken Lonsdale	KenLonsdale
19	Ken Lonsdale	KenLonsdale
20	Chris McAfee	ChrisMcAfee
21	Kean Nguyen	KeanNguyen

2.f) Now, the new generated column can be used in place of original column in original dataset table.

Steps for copying the data:

a. Click on file and select close and load.

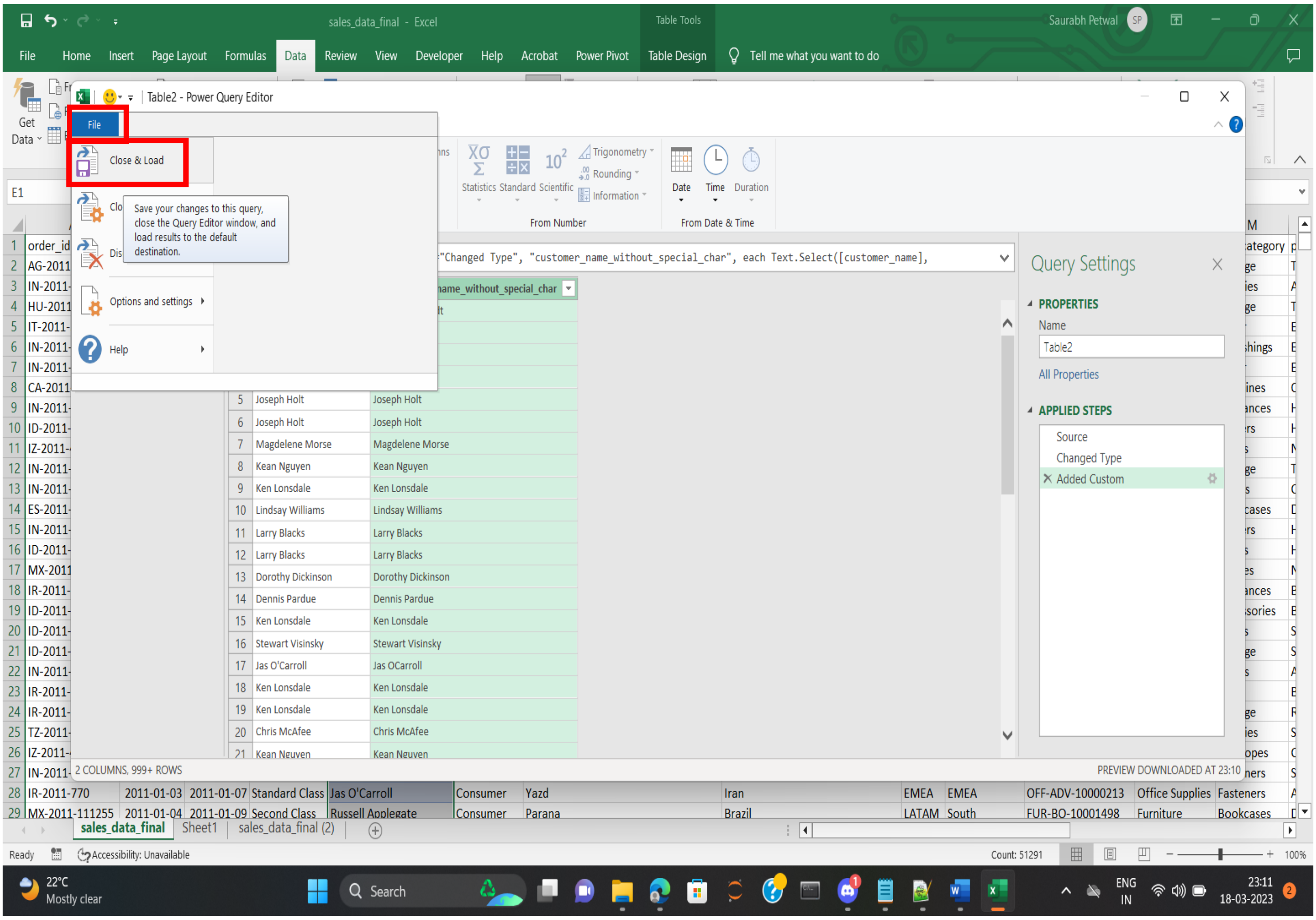
b. A separate sheet will be created on the existing workbook and from there the new cleaned column can be used in place of original uncleaned column. But in our case there are some german, latin etc characters were there so we haven't replaced new column with old existing column "customer_name" in order to save the information from getting lost, but used the new column for identifying the rows having special characters and german characters from existing column by creating a "Flag" column and comparing both existing and new column contents using "if" formula in excel as shown below:

```
=if (cell of 'customer_name_with_special_character' = cell of 'customer_name', 1, 0)
```

Then copying the formula to all rows by dragging the bottom corner of cell where the above formula is applied.

Note: The above procedure is applied to three columns to clean the data and are as follows:

1. **state**
2. **customer_name**
3. **Product_name**



Save your changes to this query, close the Query Editor window, and load results to the default destination.

```
"Changed Type", "customer_name_without_special_char", each Text.Select([customer_name],
```

		name_without_special_char
5	Joseph Holt	Joseph Holt
6	Joseph Holt	Joseph Holt
7	Magdelene Morse	Magdelene Morse
8	Kean Nguyen	Kean Nguyen
9	Ken Lonsdale	Ken Lonsdale
10	Lindsay Williams	Lindsay Williams
11	Larry Blacks	Larry Blacks
12	Larry Blacks	Larry Blacks
13	Dorothy Dickinson	Dorothy Dickinson
14	Dennis Pardue	Dennis Pardue
15	Ken Lonsdale	Ken Lonsdale
16	Stewart Visinsky	Stewart Visinsky
17	Jas O'Carroll	Jas OCarroll
18	Ken Lonsdale	Ken Lonsdale
19	Ken Lonsdale	Ken Lonsdale
20	Chris McAfee	Chris McAfee
21	Kean Nguyen	Kean Nguyen

2 COLUMNS, 999+ ROWS

PREVIEW DOWNLOADED AT 23:10

Query Settings

PROPERTIES

Name

Table2

All Properties

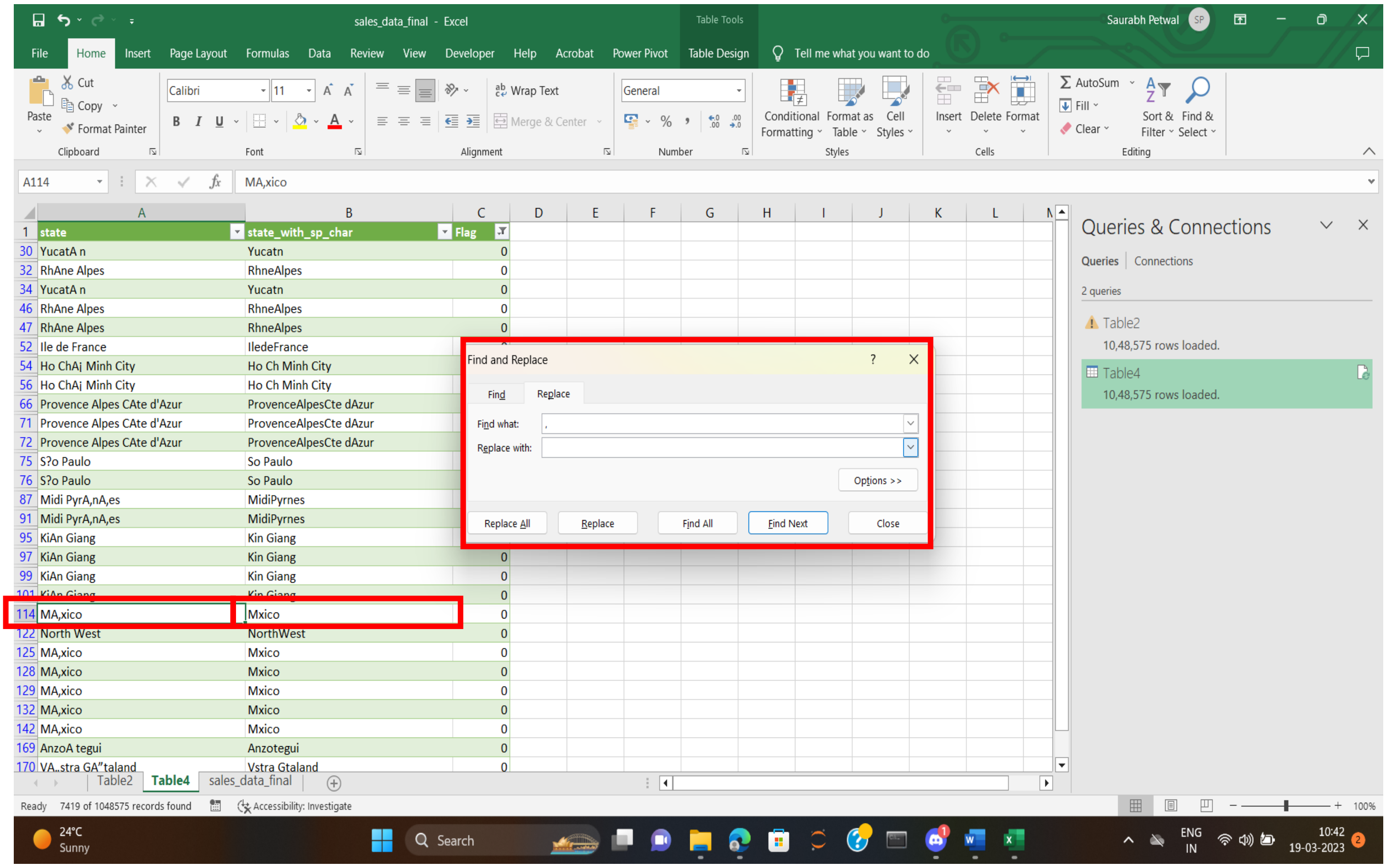
APPLIED STEPS

- Source
- Changed Type
- Added Custom

28	IR-2011-770	2011-01-03	2011-01-07	Standard Class	Jas O'Carroll	Consumer	Yazd	Iran	EMEA	EMEA	OFF-ADV-10000213	Office Supplies	Fasteners
29	MX-2011-111255	2011-01-04	2011-01-09	Second Class	Russell Applegate	Consumer	Parana	Brazil	LATAM	South	FUR-BO-10001498	Furniture	Bookcases

2.g) As from the below example we can clearly see that by using the above procedure to clean the data there might be chance of loosing information from data (characters similar to alphabets except special characters – **"MÂ,xico"** >> "**Mxico"**. So from help of flag column where the values are 0 (zero) we can check the particular columns and can refill the alphabets in the data where such Latin, German or etc words was removed.

Also UTF 8 characters were in form of space was present in the data so it was removed by replacing with "[null]" manually one by one using filter and replace function.



A114 | MA,xico

	state	state_with_sp_char	Flag
30	YucatA n	Yucatn	0
32	RhAne Alpes	RhneAlpes	0
34	YucatA n	Yucatn	0
46	RhAne Alpes	RhneAlpes	0
47	RhAne Alpes	RhneAlpes	0
52	Ile de France	IledeFrance	0
54	Ho ChAj Minh City	Ho Ch Minh City	
56	Ho ChAj Minh City	Ho Ch Minh City	
66	Provence Alpes CAte d'Azur	ProvenceAlpesCte dAzur	
71	Provence Alpes CAte d'Azur	ProvenceAlpesCte dAzur	
72	Provence Alpes CAte d'Azur	ProvenceAlpesCte dAzur	
75	S?o Paulo	So Paulo	
76	S?o Paulo	So Paulo	
87	Midi PyrA,nA,es	MidiPyrnes	
91	Midi PyrA,nA,es	MidiPyrnes	
95	KiAn Giang	Kin Giang	
97	KiAn Giang	Kin Giang	0
99	KiAn Giang	Kin Giang	0
101	KiAn Giang	Kin Giang	0
114	MA,xico	Mxico	0
122	North West	NorthWest	0
125	MA,xico	Mxico	0
128	MA,xico	Mxico	0
129	MA,xico	Mxico	0
132	MA,xico	Mxico	0
142	MA,xico	Mxico	0
169	AnzoA tegui	Anzotegui	0
170	VA..stra GA"taland	Vstra Gtaland	0

Find and Replace

Find what: ,

Replace with:

Replace All | Replace | Find All | Find Next | Close

Table2 – 10,48,575 rows loaded.

Table4 – 10,48,575 rows loaded.

7419 of 1048575 records found

---**END**---