

Data Modelling and Joins

UTKARSH GAIKWAD



Transforming the data

The screenshot shows a data transformation interface. On the left, a 'Navigator' pane lists data sources under 'Sample - Superstore 1.xlsx [6]'. The 'Returns' table is selected. On the right, a preview of the 'Returns' table is shown with two columns: 'Returned' and 'Order ID'. At the bottom right, there are three buttons: 'Load', 'Transform Data', and 'Cancel'. The 'Transform Data' button is highlighted with a red rectangle. An arrow points from a text box to this button.

Navigator

Display Options ▾

- Sample - Superstore 1.xlsx [6]
 - ☒ Orders
 - ☐ People
 - ☒ Returns
 - ☐ Orders1
 - ☐ People2
 - ☐ Returns3

Returns

Column1	Column2
Returned	Order ID
Yes	CA-2019-100762
Yes	CA-2019-100762
Yes	CA-2019-100762
Yes	CA-2019-100762
Yes	CA-2019-100867
Yes	CA-2019-102652
Yes	CA-2019-102652
Yes	CA-2019-102652
Yes	CA-2019-102652
Yes	CA-2019-103373
Yes	CA-2019-103744
Yes	CA-2019-103744
Yes	CA-2019-103940
Yes	CA-2019-103940
Yes	CA-2019-103940
Yes	CA-2019-103940
Yes	CA-2019-103940
Yes	CA-2019-104829
Yes	CA-2019-105270
Yes	CA-2019-105270
Yes	CA-2019-108609
Yes	CA-2019-108861
Yes	CA-2019-109918
Yes	CA-2019-109918

Load Transform Data Cancel

Click to Transform Data

Changing First Column to Header

Step 3: Use first row header

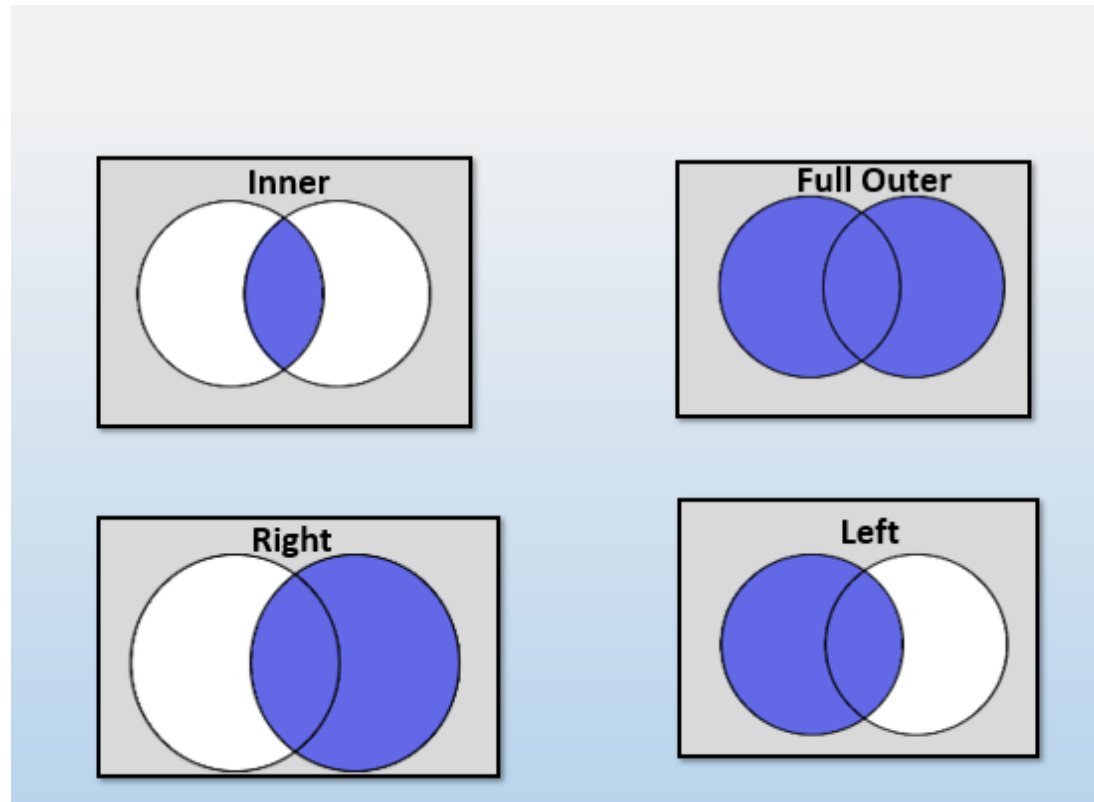
Step 2: Click Transform

The screenshot shows the Power Query Editor interface. On the left, the 'Queries' pane lists 'Orders' and 'Returns', with 'Returns' selected. The main area displays the 'Transform' tab of the ribbon, where the 'Use First Row as Headers' option is highlighted. The data table has two columns: 'Returned' and 'Order ID'. The 'Returned' column contains values like 'Yes' and 'No', and the 'Order ID' column contains unique identifiers. The 'Query Settings' pane on the right shows the 'Properties' and 'Applied Steps' sections.

	Returned	Order ID
1	Yes	CA-2019-100762
2	Yes	CA-2019-100762
3	Yes	CA-2019-100762
4	Yes	CA-2019-100762
5	Yes	CA-2019-100867
6	Yes	CA-2019-102652
7	Yes	CA-2019-102652
8	Yes	CA-2019-102652
9	Yes	CA-2019-102652
10	Yes	CA-2019-103373
11	Yes	CA-2019-103744
12	Yes	CA-2019-103744
13	Yes	CA-2019-103940
14	Yes	CA-2019-103940
15	Yes	CA-2019-103940
16	Yes	CA-2019-103940
17	Yes	CA-2019-104829
18	Yes	CA-2019-105270
19	Yes	CA-2019-105270
20	Yes	CA-2019-108609
21	Yes	CA-2019-108861
22	Yes	CA-2019-109918
23	Yes	CA-2019-109918
24	Yes	CA-2019-109918
25	Yes	CA-2019-110786
26	Yes	CA-2019-110786
27	Yes	CA-2019-110786
28	Yes	CA-2019-110786

Step 1: Click Returns

Joins In power BI



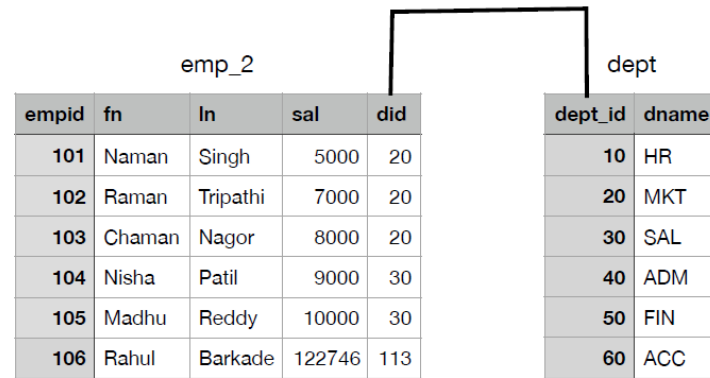
Inner Join

emp_2					dept	
empid	fn	ln	sal	did	dept_id	dname
101	Naman	Singh	5000	20	10	HR
102	Raman	Tripathi	7000	20	20	MKT
103	Chaman	Nagor	8000	20	30	SAL
104	Nisha	Patil	9000	30	40	ADM
105	Madhu	Reddy	10000	30	50	FIN
106	Rahul	Barkade	122746	113	60	ACC

INNER_JOIN_OP

empid	fn	ln	sal	did	dept_id	dname
101	Naman	Singh	5000	20	20	MKT
102	Raman	Tripathi	7000	20	20	MKT
103	Chaman	Nagor	8000	20	20	MKT
104	Nisha	Patil	9000	30	30	SAL
105	Madhu	Reddy	10000	30	30	SAL

Left Join



LO_JOIN_OP

empid	fn	ln	sal	did	dept_id	dname
101	Naman	Singh	5000	20	20	MKT
102	Raman	Tripathi	7000	20	20	MKT
103	Chaman	Nagor	8000	20	20	MKT
104	Nisha	Patil	9000	30	30	SAL
105	Madhu	Reddy	10000	30	30	SAL
106	Rahul	Barkade	122746	113	NULL	NULL

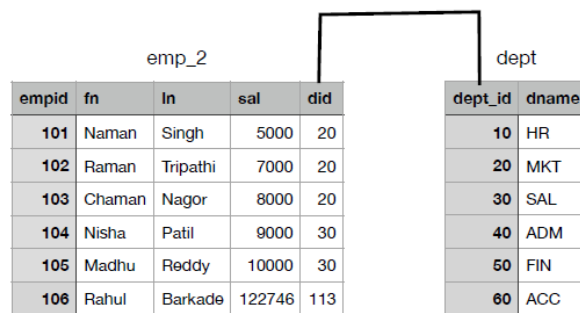
Right Join

emp_2					dept	
empid	fn	ln	sal	did	dept_id	dname
101	Naman	Singh	5000	20	10	HR
102	Raman	Tripathi	7000	20	20	MKT
103	Chaman	Nagor	8000	20	30	SAL
104	Nisha	Patil	9000	30	40	ADM
105	Madhu	Reddy	10000	30	50	FIN
106	Rahul	Barkade	122746	113	60	ACC

RO_JOIN_OP

empid	fn	ln	sal	did	dept_id	dname
NULL	NULL	NULL	NULL	NULL	10	HR
102	Raman	Tripathi	7000	20	20	MKT
103	Chaman	Nagor	8000	20	20	MKT
101	Naman	Singh	5000	20	20	MKT
104	Nisha	Patil	9000	30	30	SAL
105	Madhu	Reddy	10000	30	30	SAL
NULL	NULL	NULL	NULL	NULL	40	ADM
NULL	NULL	NULL	NULL	NULL	50	FIN
NULL	NULL	NULL	NULL	NULL	60	ACC

Full Outer



FO_JOIN_OP

empid	fn	ln	sal	did	dept_id	dname
NULL	NULL	NULL	NULL	NULL	10	HR
102	Raman	Tripathi	7000	20	20	MKT
103	Chaman	Nagor	8000	20	20	MKT
101	Naman	Singh	5000	20	20	MKT
104	Nisha	Patil	9000	30	30	SAL
105	Madhu	Reddy	10000	30	30	SAL
NULL	NULL	NULL	NULL	NULL	40	ADM
NULL	NULL	NULL	NULL	NULL	50	FIN
NULL	NULL	NULL	NULL	NULL	60	ACC
106	Rahul	Barkade	122746	113	NULL	NULL

Creating Joins In PowerBI

The screenshot shows the Power Query Editor interface. The ribbon at the top includes tabs for File, Home, Transform, Add Column, View, Tools, and Help. The 'Merge Queries' button is highlighted with a red box. A text box with the text 'Click Merge Queries' has an arrow pointing to this button. The main area displays a table with the following columns: Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, and Customer Name. The table contains 27 rows of data. The status bar at the bottom indicates '21 COLUMNS, 999+ ROWS' and 'Column profiling based on top 1000 rows'. The system tray at the very bottom shows the date and time as '12:54 AM 04-10-2023'.

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name
1	CA-2021-152156	08-11-2021	11-11-2021	Second Class	CG-12520	Claire Gute
2	CA-2021-152156	08-11-2021	11-11-2021	Second Class	CG-12520	Claire Gute
3	CA-2021-138688	12-06-2021	16-06-2021	Second Class	DV-13045	Darrin Van H
4	US-2020-108966	11-10-2020	18-10-2020	Standard Class	SO-20335	Sean O'Donn
5	US-2020-108966	11-10-2020	18-10-2020	Standard Class	SO-20335	Sean O'Donn
6	CA-2019-115812	09-06-2019	14-06-2019	Standard Class	BH-11710	Brosina Hoffr
7	CA-2019-115812	09-06-2019	14-06-2019	Standard Class	BH-11710	Brosina Hoffr
8	CA-2019-115812	09-06-2019	14-06-2019	Standard Class	BH-11710	Brosina Hoffr
9	CA-2019-115812	09-06-2019	14-06-2019	Standard Class	BH-11710	Brosina Hoffr
10	CA-2019-115812	09-06-2019	14-06-2019	Standard Class	BH-11710	Brosina Hoffr
11	CA-2019-115812	09-06-2019	14-06-2019	Standard Class	BH-11710	Brosina Hoffr
12	CA-2019-115812	09-06-2019	14-06-2019	Standard Class	BH-11710	Brosina Hoffr
13	CA-2022-114412	15-04-2022	20-04-2022	Standard Class	AA-10480	Andrew Allen
14	CA-2021-161389	05-12-2021	10-12-2021	Standard Class	IM-15070	Irene Maddox
15	US-2020-118983	22-11-2020	26-11-2020	Standard Class	HP-14815	Harold Pawla
16	US-2020-118983	22-11-2020	26-11-2020	Standard Class	HP-14815	Harold Pawla
17	CA-2019-105893	11-11-2019	18-11-2019	Standard Class	PK-19075	Pete Kriz
18	CA-2019-167164	13-05-2019	15-05-2019	Second Class	AG-10270	Alejandro Grc
19	CA-2019-143336	27-08-2019	01-09-2019	Second Class	ZD-21925	Zuschuss Don
20	CA-2019-143336	27-08-2019	01-09-2019	Second Class	ZD-21925	Zuschuss Don
21	CA-2019-143336	27-08-2019	01-09-2019	Second Class	ZD-21925	Zuschuss Don
22	CA-2021-137330	09-12-2021	13-12-2021	Standard Class	KB-16585	Ken Black
23	CA-2021-137330	09-12-2021	13-12-2021	Standard Class	KB-16585	Ken Black
24	US-2022-156909	16-07-2022	18-07-2022	Second Class	SF-20065	Sandra Flana
25	CA-2020-106320	25-09-2020	30-09-2020	Standard Class	EB-13870	Emily Burns
26	CA-2021-121755	16-01-2021	20-01-2021	Second Class	EH-13945	Eric Hoffmani
27	CA-2021-121755	16-01-2021	20-01-2021	Second Class	EH-13945	Eric Hoffmani

Creating Joins In PowerBI Contd.

The screenshot shows the 'Merge' dialog box in PowerBI. It contains two tables: 'Orders' and 'Returns'. The 'Orders' table has columns: Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, and Segment. The 'Returns' table has columns: Returned, Order ID, and a list of returned items. The 'Join Kind' is set to 'Left Outer (all from first, matching from second)'. The 'OK' button is highlighted. Four callout boxes provide instructions: Step 1: Select Returns (points to the 'Returns' dropdown), Step 2: Join Type (points to the 'Join Kind' dropdown), Step 3: Ctrl Click And Select Both Columns (points to the 'Order ID' columns in both tables), and Step 4: Click Ok (points to the 'OK' button).

Step 1: Select Returns

Step 2: Join Type

Step 3: Ctrl Click And Select Both Columns

Step 4: Click Ok

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment
1	CA-2021-152156	08-11-2021	11-11-2021	Second Class	CG-12520	Claire Gute	Consumer
2	CA-2021-152156	08-11-2021	11-11-2021	Second Class	CG-12520	Claire Gute	Consumer
3	CA-2021-138688	12-06-2021	16-06-2021	Second Class	DV-13045	Darrin Van Huff	Corporate
4	US-2020-108966	11-10-2020	18-10-2020	Standard Class	SO-20335	Sean O'Donnell	Consumer

Returned	Order ID
Yes	CA-2019-100762
Yes	CA-2019-100762
Yes	CA-2019-100762
Yes	CA-2019-100762
Yes	CA-2019-100867

Join Kind: Left Outer (all from first, matching from second)

☐ Use fuzzy matching to perform the merge

► Fuzzy matching options

✓ The selection matches 800 of 9994 rows from the first table.

OK Cancel

Expand Returns Table

Power Query Editor interface showing a table with columns: Item, Sales, Quantity, Discount, Profit, and Returns. The Returns column contains table names. A red box highlights the expand icon (two arrows pointing right) in the top right corner of the Returns column header. An arrow points from a text box 'Click Expand' to this icon.

Item	Sales	Quantity	Discount	Profit	Returns
on Bookcase	261.96	2	0	41.9136	Table
olstered Stacking Chairs, Rounded Back	731.94	3	0	219.582	Table
Labels for Typewriters by Universal	14.62	2	0	6.8714	Table
Slim Rectangular Table	957.5775	5	0.45	-383.031	Table
System	22.368	2	0.2	2.5164	Table
d and Plastic Desk Accessories, Cherry Wood	48.86	7	0	14.1694	Table
	7.28	4	0	1.9656	Table
MP phone	907.152	6	0.2	90.7152	Table
s with Locking Rings by Samsill	18.504	3	0.2	5.7825	Table
utlet Surge	114.9	5	0	34.47	Table
r Conference Tables	1706.184	9	0.2	85.3092	Table
phone - Charcoal black	911.424	4	0.2	68.3568	Table
	15.552	3	0.2	5.4432	Table
Comb Binding Machine	407.976	3	0.2	132.5922	Table
lter for HEPA Air Cleaner, Very Large Room, HE...	68.81	5	0.8	-123.858	Table
led Plastic Frosted Binders	2.544	3	0.8	-3.816	Table
ertical 5-Shelf: 72"H x 36"W x 18 1/2"D	665.88	6	0	13.3176	Table
rawer	55.5	2	0	9.99	Table
	8.56	2	0	2.4824	Table
ve	213.48	3	0.2	16.011	Table
ew Binder, White, 1"	22.72	4	0.2	7.384	Table
	19.46	7	0	5.0596	Table
Strip, 4' Cord Length	60.34	7	0	15.6884	Table
Chair, Gray	71.372	2	0.3	-1.0196	Table
Slim Rectangular Table	1044.63	3	0	240.2649	Table
e Binders	11.648	2	0.2	4.2224	Table
Drive USB 2.0 Flash Drive	90.57	3	0	11.7741	Table

Replacing null values with no in Returns column

The screenshot shows the Power Query Editor interface. The 'Transform' tab is active, and the 'Replace Values' button is highlighted. A dialog box is open, prompting the user to replace 'null' values with 'No' in the selected columns. The 'Returns' column is selected in the data table. The 'OK' button is highlighted in the dialog box.

Numbered callouts:

- 1: Points to the 'Returns' column header in the data table.
- 2: Points to the 'Transform' tab in the ribbon.
- 3: Points to the 'Replace Values' button in the ribbon.
- 4: Points to the 'Value To Find' field in the 'Replace Values' dialog box.
- 5: Points to the 'OK' button in the 'Replace Values' dialog box.

Applying all changes

The screenshot displays the Microsoft Power Query Editor window. The ribbon at the top includes tabs for File, Home, Transform, Add Column, View, Tools, and Help. The 'File' tab is currently selected, showing options like Close & Apply, New Source, Recent Sources, Enter Data, Data source settings, Manage Parameters, Refresh Preview, Advanced Editor, Choose Columns, Remove Columns, Keep Rows, Remove Rows, Split Column, Group By, Data Type: Text, Use First Row as Headers, Replace Values, Merge Queries, Append Queries, Combine Files, Text Analytics, Vision, and Azure Machine Learning. The main area shows a table with columns: 1.2 Sales, 1.2 Quantity, 1.2 Discount, 1.2 Profit, and Returns.Returned. The table contains 28 rows of data. The 'Applied Steps' pane on the right shows a list of steps: Source, Navigation, Promoted Headers, Changed Type, Merged Queries, Expanded Returns, and Replaced Value. The 'Replaced Value' step is selected.

	1.2 Sales	1.2 Quantity	1.2 Discount	1.2 Profit	Returns.Returned
1	Bookcase	261.96	2	0	41.9136 No
2	ered Stacking Chairs, Rounded Back	731.94	3	0	219.582 No
3	als for Typewriters by Universal	14.62	2	0	6.8714 No
4	m Rectangular Table	957.5775	5	0.45	-383.031 No
5	em	22.368	2	0.2	2.5164 No
6	nd Plastic Desk Accessories, Cherry Wood	48.86	7	0	14.1694 No
7		7.28	4	0	1.9656 No
8	ophone	907.152	6	0.2	90.7152 No
9	th Locking Rings by Samsill	18.504	3	0.2	5.7825 No
10	t Surge	114.9	5	0	34.47 No
11	nference Tables	1706.184	9	0.2	85.3092 No
12	ione - Charcoal black	911.424	4	0.2	68.3568 No
13		15.552	3	0.2	5.4432 No
14	mb Binding Machine	407.976	3	0.2	132.5922 No
15	r for HEPA Air Cleaner, Very Large Room, HE...	68.81	5	0.8	-123.858 No
16	Plastic Frosted Binders	2.544	3	0.8	-3.816 No
17	cal 5-Shelf: 72"H x 36"W x 18 1/2"D	665.88	6	0	13.3176 No
18	er	55.5	2	0	9.99 No
19		8.56	2	0	2.4824 Yes
20		8.56	2	0	2.4824 Yes
21		8.56	2	0	2.4824 Yes
22		213.48	3	0.2	16.011 Yes
23		213.48	3	0.2	16.011 Yes
24		213.48	3	0.2	16.011 Yes
25	v Binder, White, 1"	22.72	4	0.2	7.384 Yes
26	v Binder, White, 1"	22.72	4	0.2	7.384 Yes
27	v Binder, White, 1"	22.72	4	0.2	7.384 Yes
28					

23 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

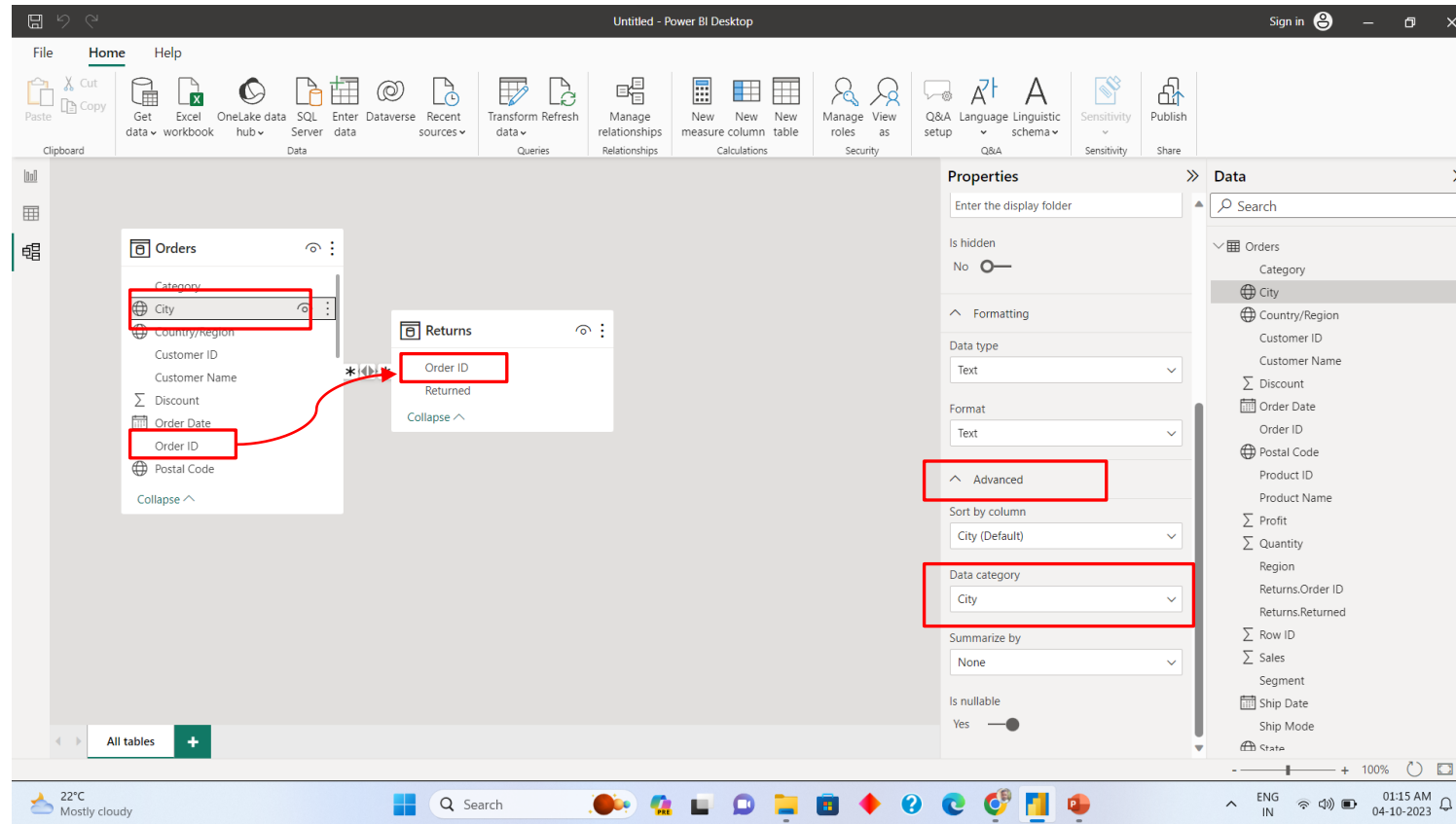
PREVIEW DOWNLOADED AT 12:13 AM

22°C Mostly cloudy

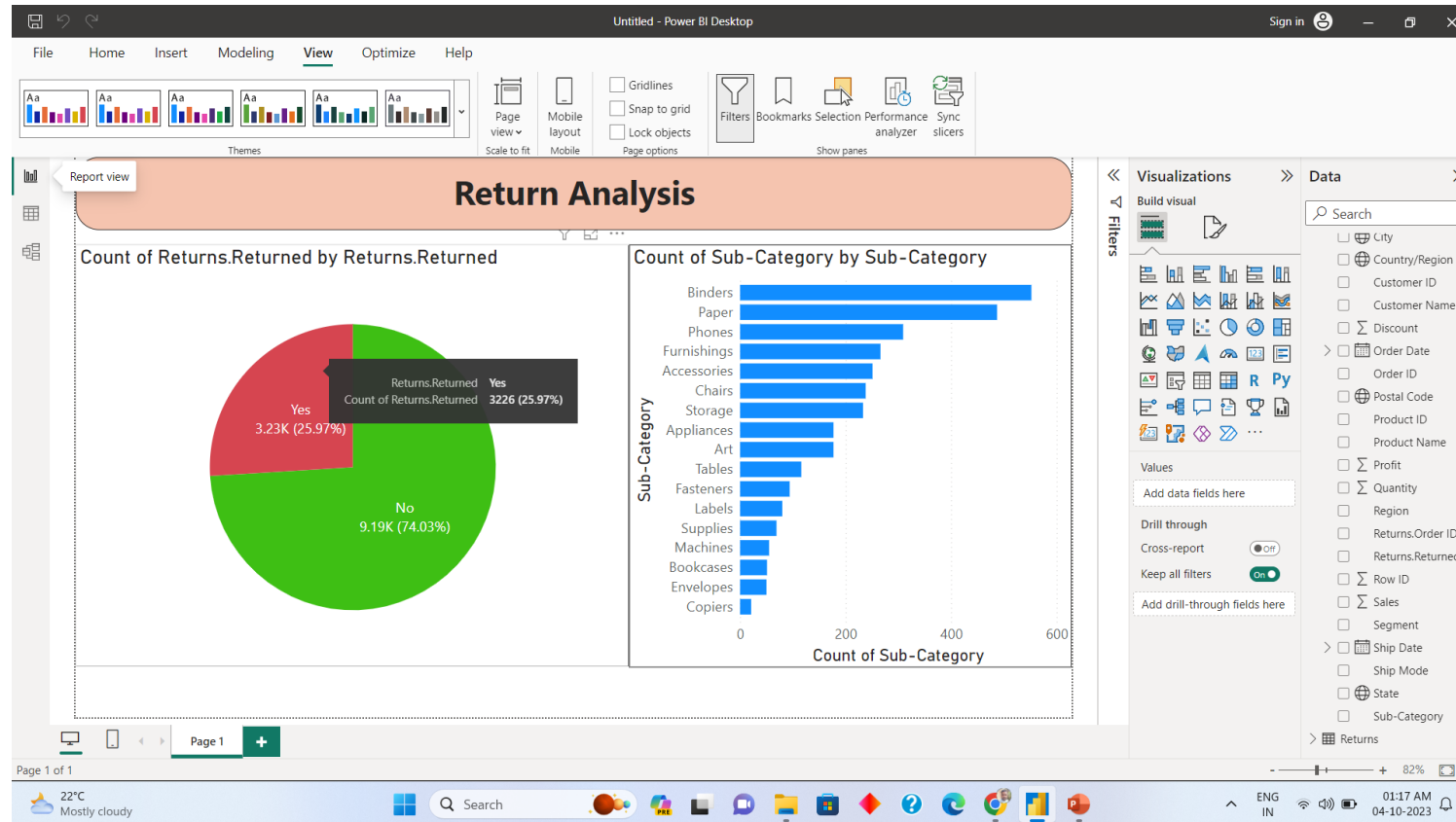
Search

ENG IN 01:06 AM 04-10-2023

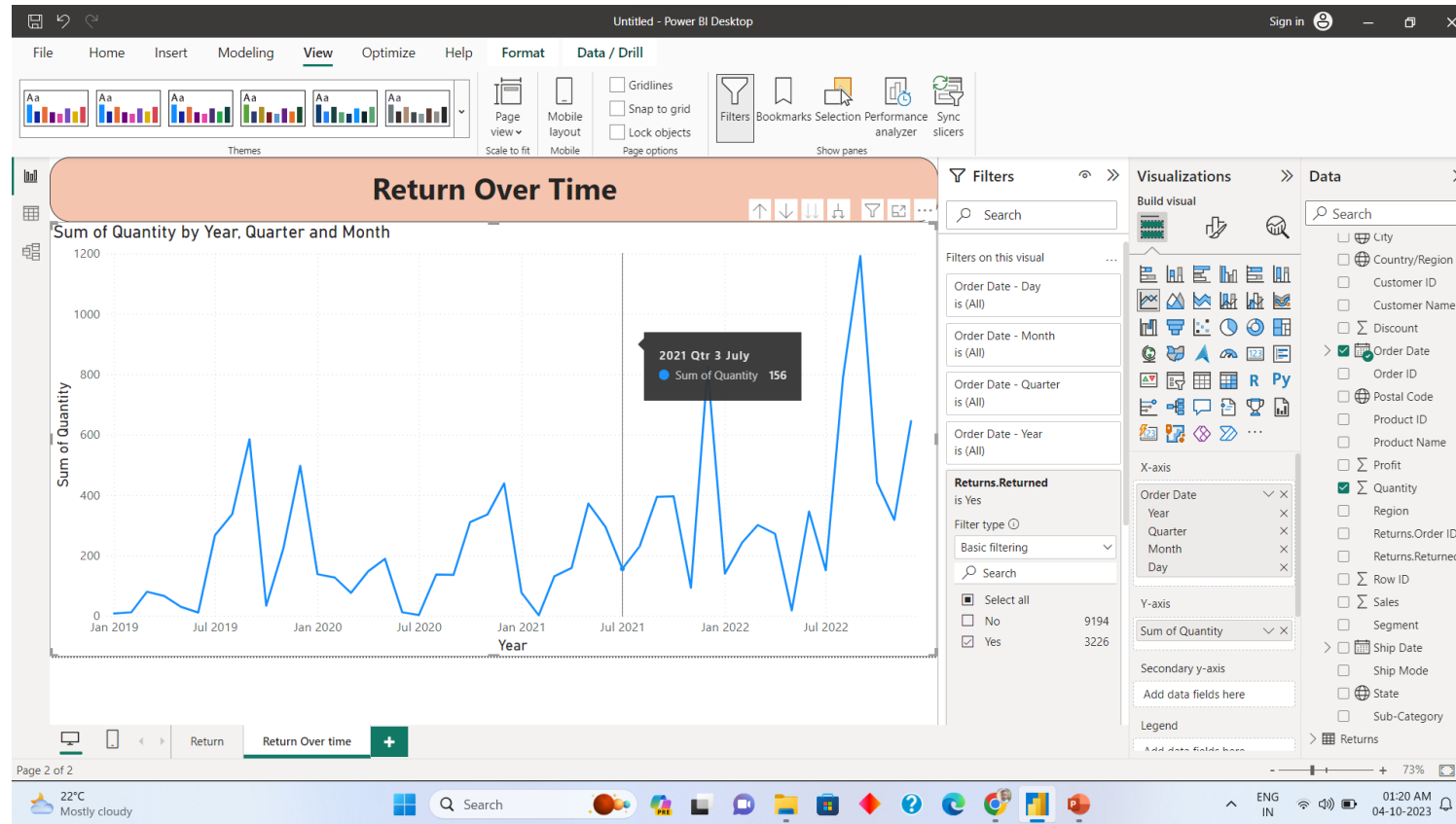
Data Modelling



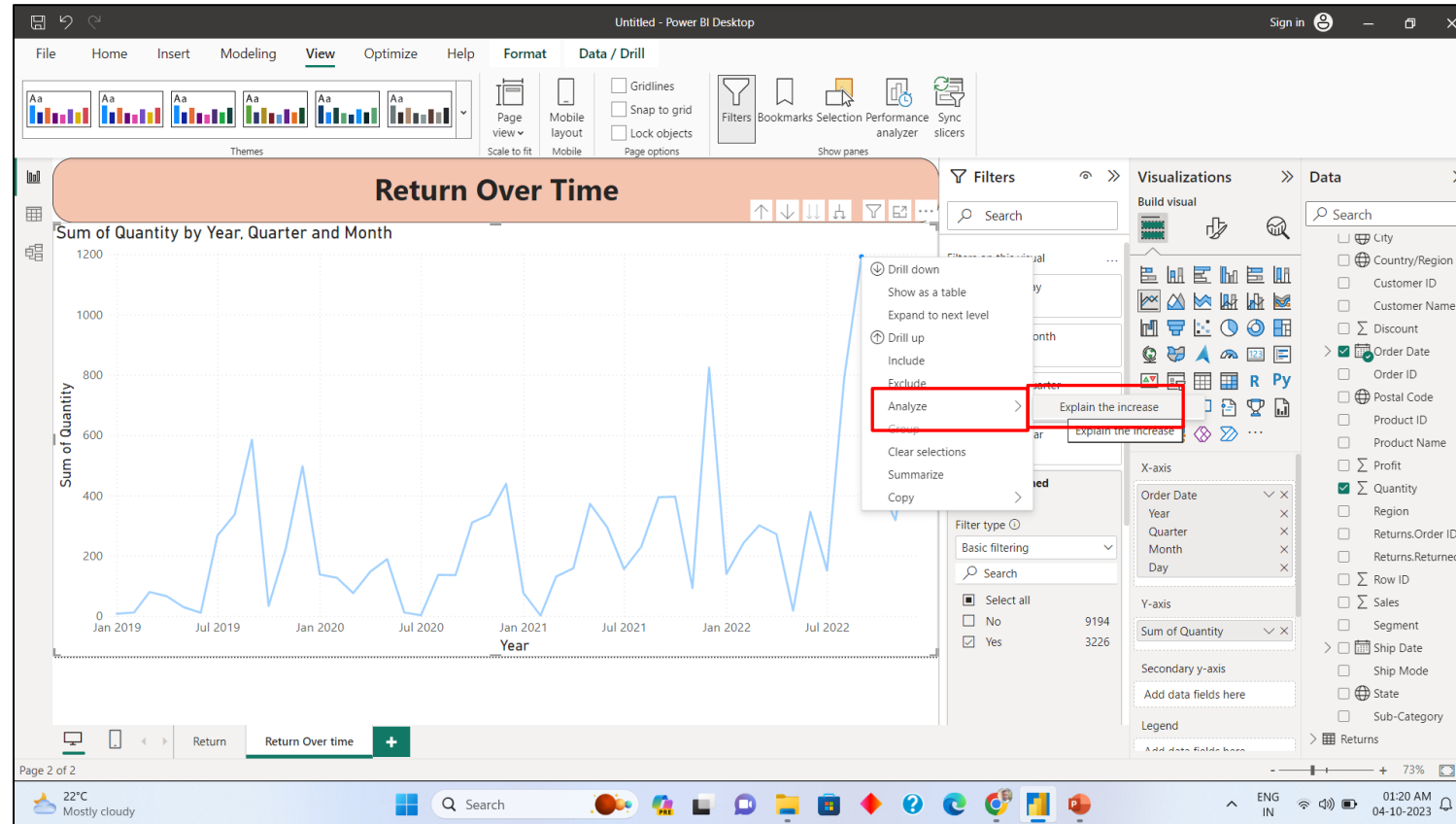
Return Analysis



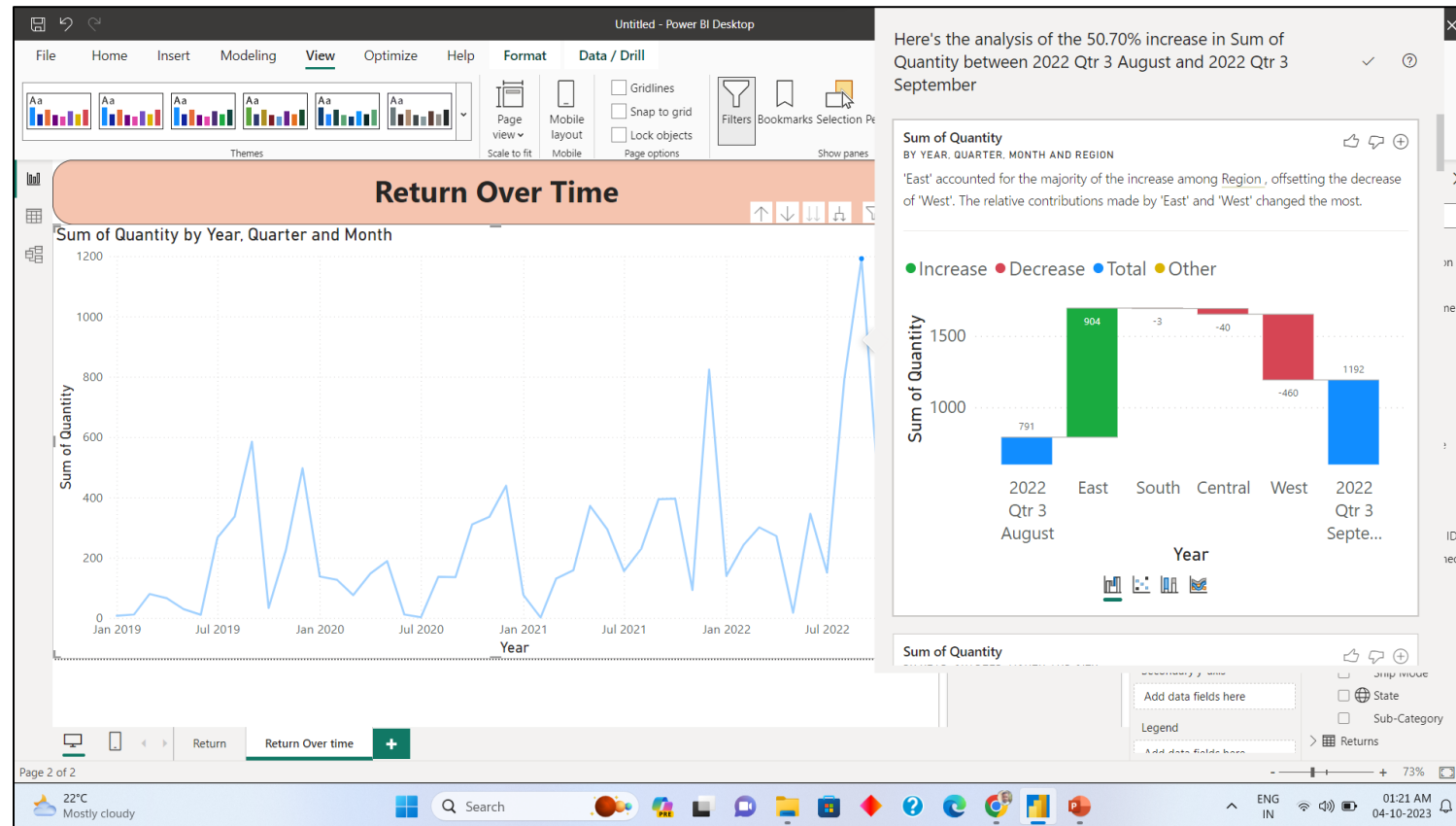
Return Quantity over Time



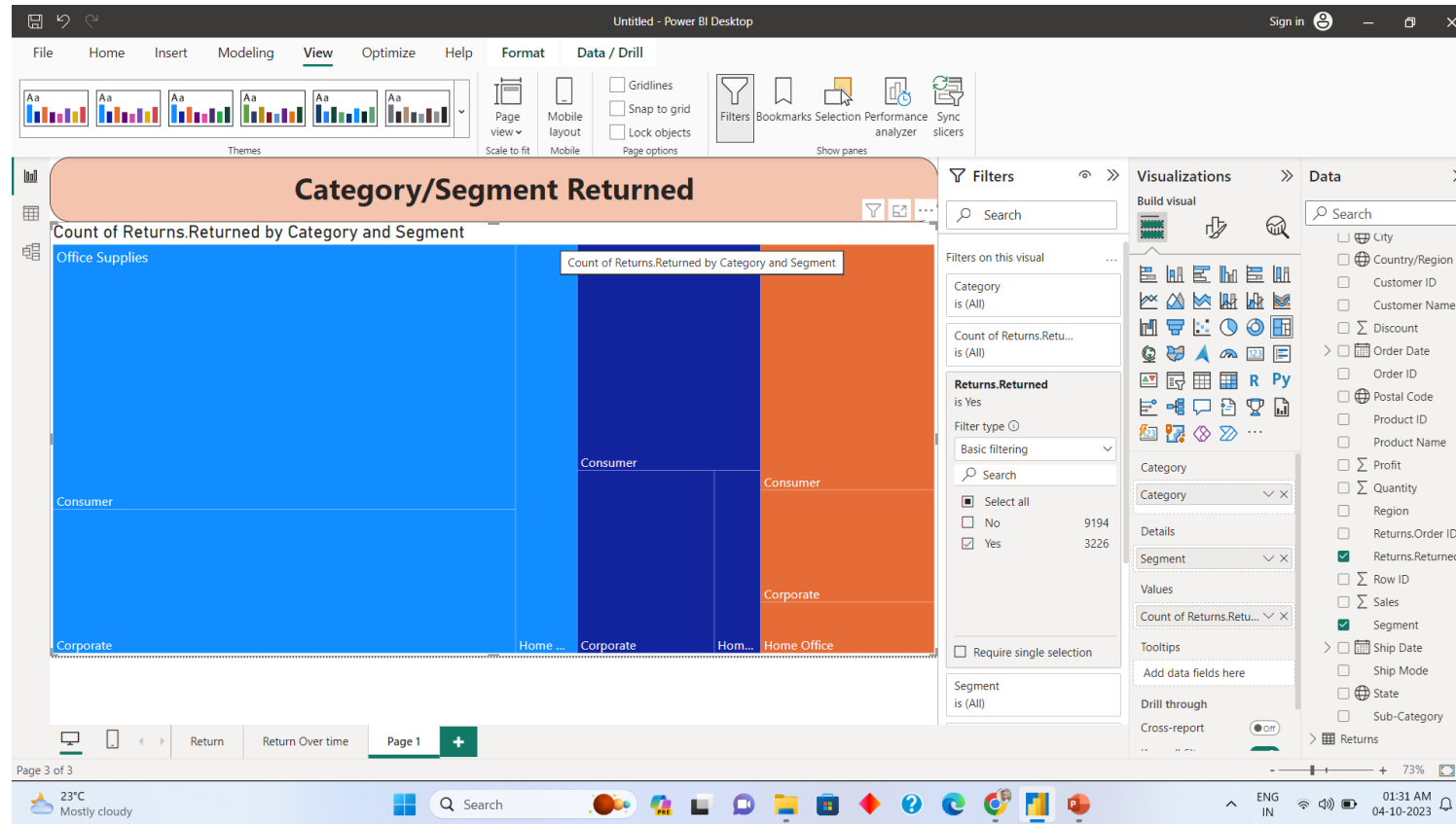
Understanding why Return Quantity increased



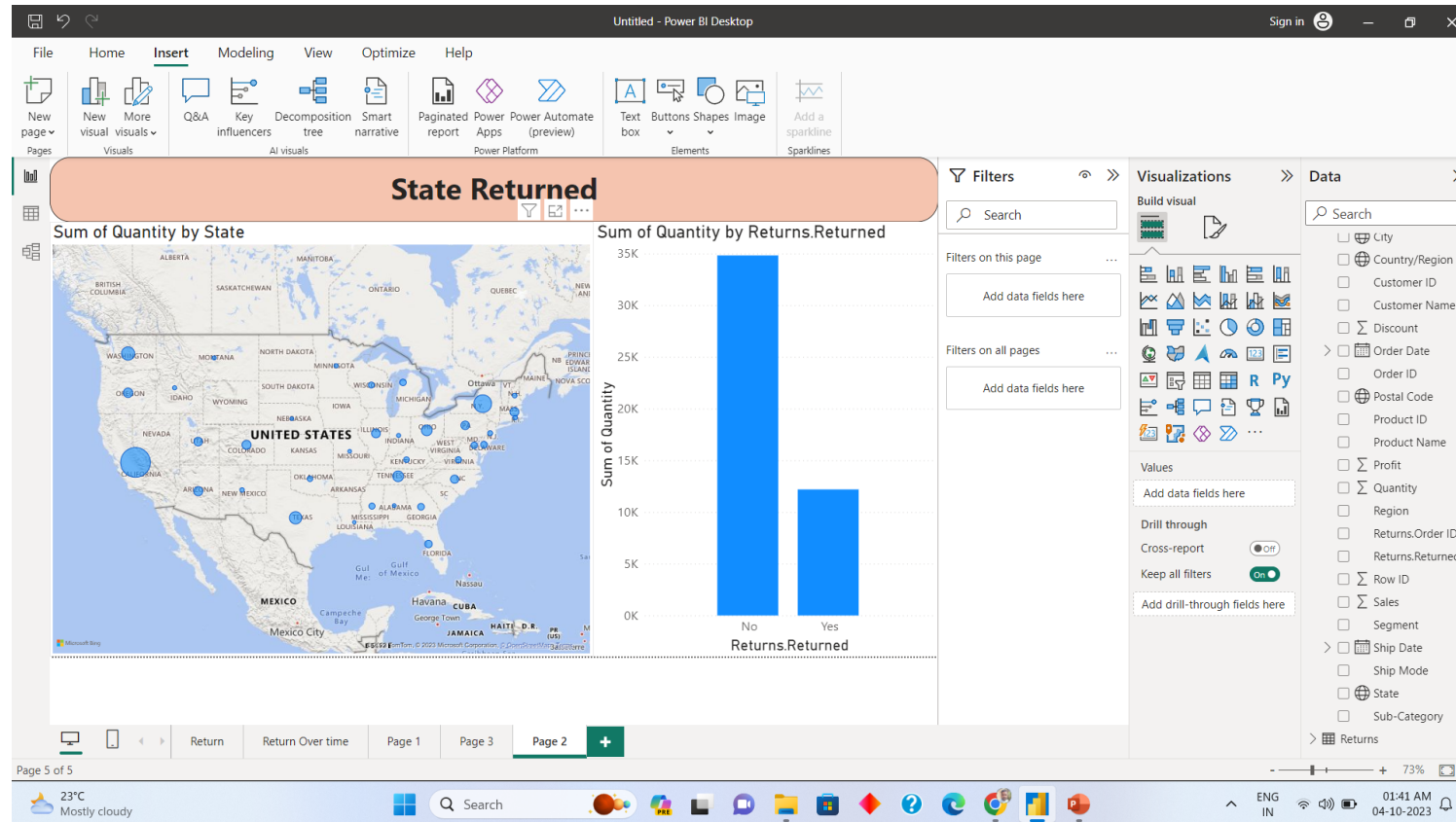
Analysis Results



Category/Segment Returns



Quantity Returned by State



Thank you

UTKARSH GAIKWAD

PING ME ON SKYPE FOR ANY QUERIES