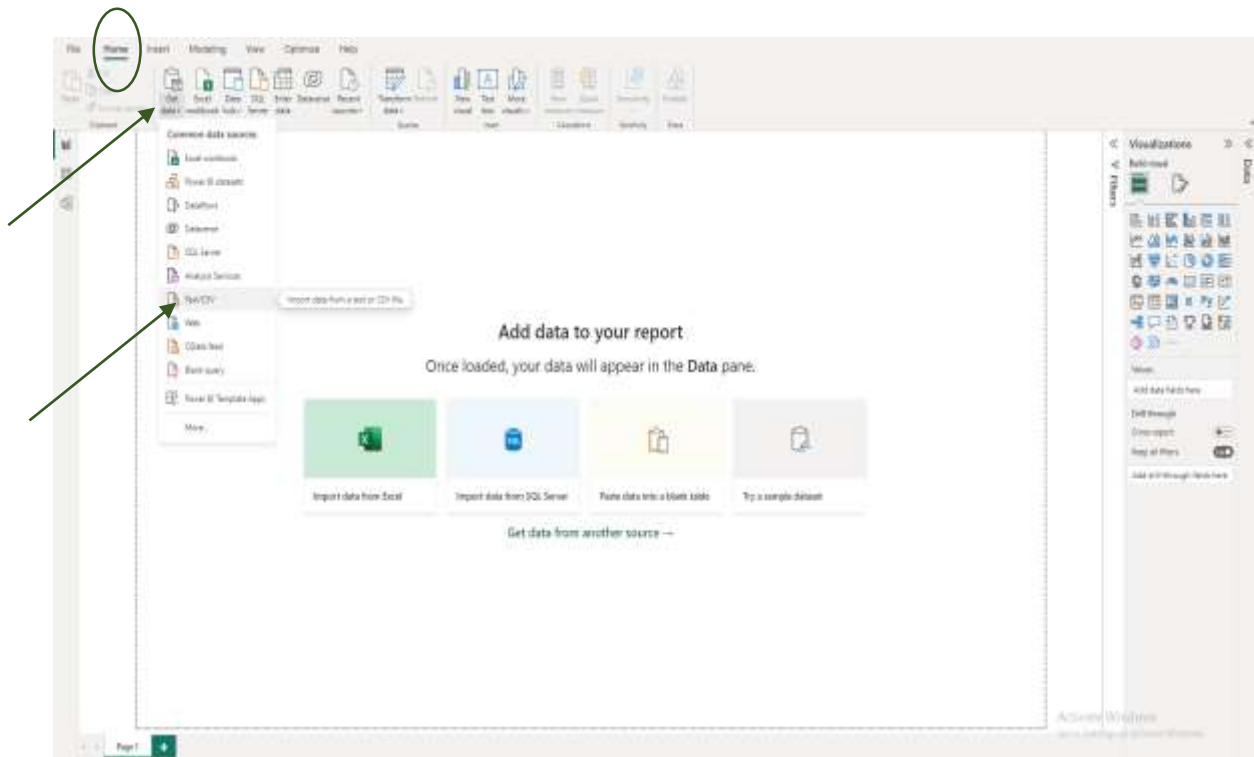
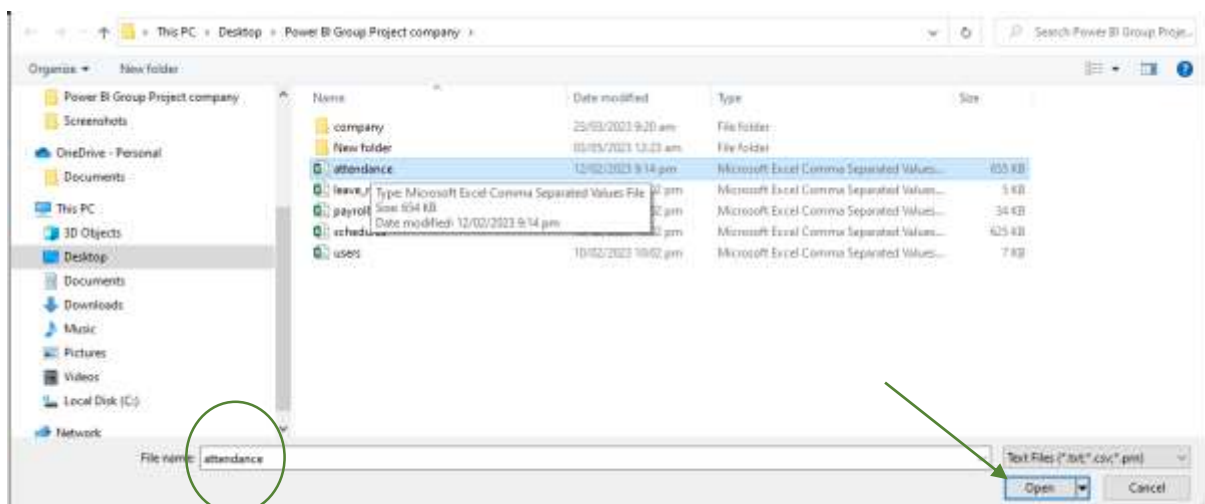


Methodology

Importing Data to Power BI. Open Power BI Desktop on your PC. Go to Home and click Get Data then select Text/CSV.



Find the file you downloaded in your PC. Choose the attendance data to be import and click open.



This new window will pop up to the screen. This window is the preview of the file you are about to import. Press Transform Data first, because we still need to clean and transform the data.

attendance.csv

File Origin

1252: Western European (Windows)

Delimiter

Comma

Data Type Detection

Based on first 200 rows

user_id	first_name	last_name	location	date	time	timezone	cases	source
74025				20/10/2021	8:08:45 pm	+08:00	IN	mobile
74025				20/10/2021	8:09:11 pm	+08:00	IN	mobile
74027				20/10/2021	10:29:56 pm	+08:00	IN	mobile
74027				21/10/2021	12:15:33 am	+08:00	OUT	mobile
74025			Nu Orange	21/10/2021	1:22:20 pm	+08:00	IN	mobile
74027			Nu Orange	21/10/2021	3:47:52 pm	+08:00	IN	mobile
74135			Nu Orange	21/10/2021	6:09:37 pm	+08:00	IN	mobile
74025			Nu Orange	22/10/2021	1:40:57 pm	+08:00	IN	mobile
74138				22/10/2021	2:33:25 pm	+08:00	IN	mobile
74027			Nu Orange	22/10/2021	4:15:48 pm	+08:00	IN	mobile
74050			Clinic	25/10/2021	1:51:35 pm	+08:00	IN	mobile
74027				25/10/2021	1:58:21 pm	+08:00	IN	mobile
74053			Clinic	25/10/2021	2:20:04 pm	+08:00	IN	mobile
74049			Clinic	25/10/2021	2:20:16 pm	+08:00	IN	mobile
74050				25/10/2021	2:20:21 pm	+08:00	IN	mobile
74465			Nu Orange	25/10/2021	2:20:25 pm	+08:00	IN	mobile
74054			Nu Orange	25/10/2021	2:20:29 pm	+08:00	IN	mobile
74556			Nu Orange	25/10/2021	2:20:29 pm	+08:00	IN	mobile
74053			Clinic	25/10/2021	2:20:44 pm	+08:00	IN	mobile
74054			Nu Orange	25/10/2021	2:21:08 pm	+08:00	OUT	mobile

The data in the preview has been truncated due to size limits.

Extract Table Using Examples

Load

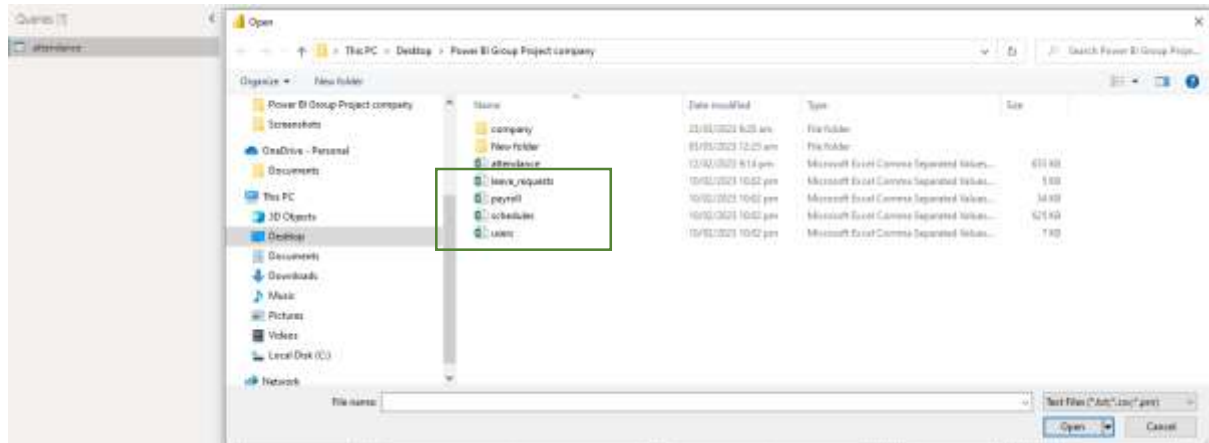
Transform Data

Cancel

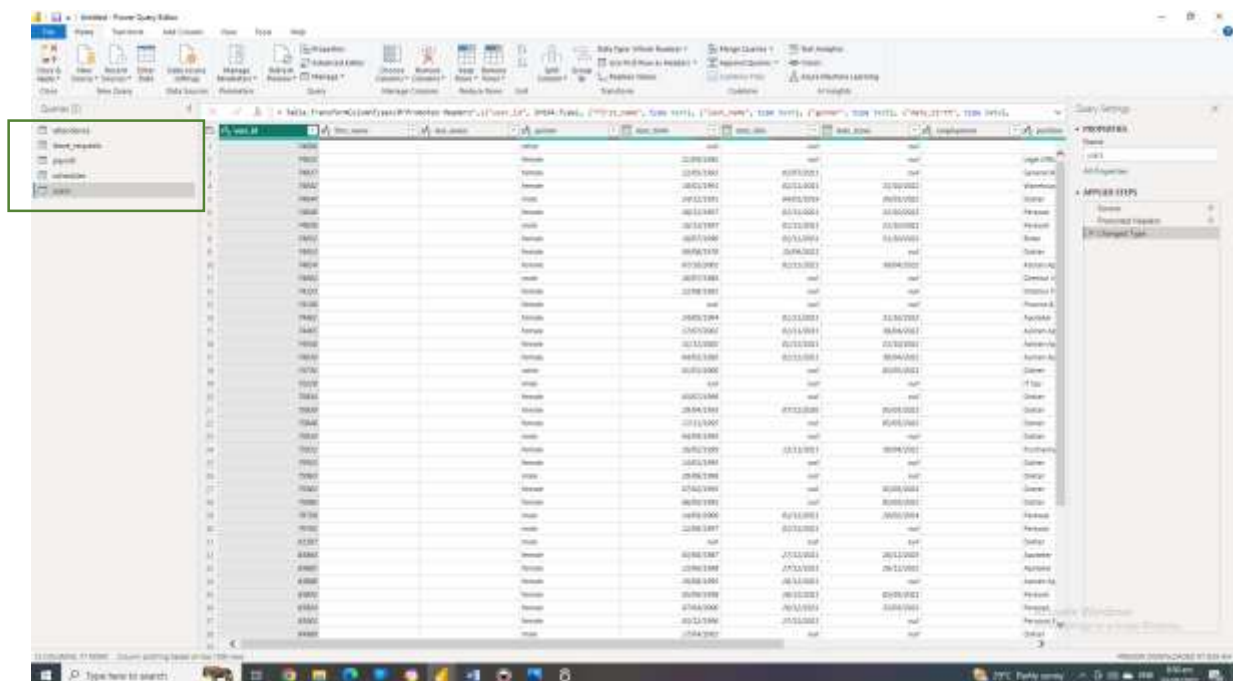
This is the final result of the steps we took to import the data.

user_id	first_name	last_name	location	date	time	timezone	cases	source
74025				20/10/2021	8:08:45 pm	+08:00	IN	mobile
74025				20/10/2021	8:09:11 pm	+08:00	IN	mobile
74027				20/10/2021	10:29:56 pm	+08:00	IN	mobile
74027				21/10/2021	12:15:33 am	+08:00	OUT	mobile
74025			Nu Orange	21/10/2021	1:22:20 pm	+08:00	IN	mobile
74027			Nu Orange	21/10/2021	3:47:52 pm	+08:00	IN	mobile
74135			Nu Orange	21/10/2021	6:09:37 pm	+08:00	IN	mobile
74025			Nu Orange	22/10/2021	1:40:57 pm	+08:00	IN	mobile
74138				22/10/2021	2:33:25 pm	+08:00	IN	mobile
74027			Nu Orange	22/10/2021	4:15:48 pm	+08:00	IN	mobile
74050			Clinic	25/10/2021	1:51:35 pm	+08:00	IN	mobile
74027				25/10/2021	1:58:21 pm	+08:00	IN	mobile
74053			Clinic	25/10/2021	2:20:04 pm	+08:00	IN	mobile
74049			Clinic	25/10/2021	2:20:16 pm	+08:00	IN	mobile
74050				25/10/2021	2:20:21 pm	+08:00	IN	mobile
74465			Nu Orange	25/10/2021	2:20:25 pm	+08:00	IN	mobile
74054			Nu Orange	25/10/2021	2:20:29 pm	+08:00	IN	mobile
74556			Nu Orange	25/10/2021	2:20:29 pm	+08:00	IN	mobile
74053			Clinic	25/10/2021	2:20:44 pm	+08:00	IN	mobile
74054			Nu Orange	25/10/2021	2:21:08 pm	+08:00	OUT	mobile

Repeat the same process we did for importing attendance to import the schedules, payroll, leave_requests and users files.

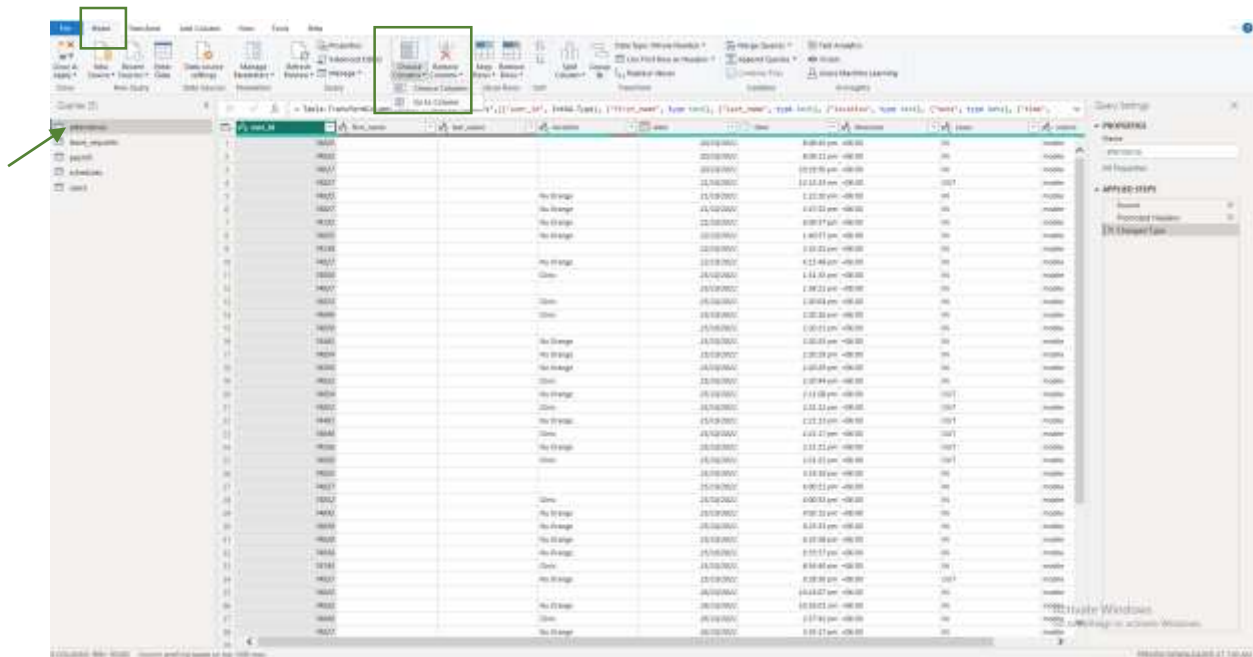


After all the files are imported into Power BI. The next step is the cleaning and transformation of the data.

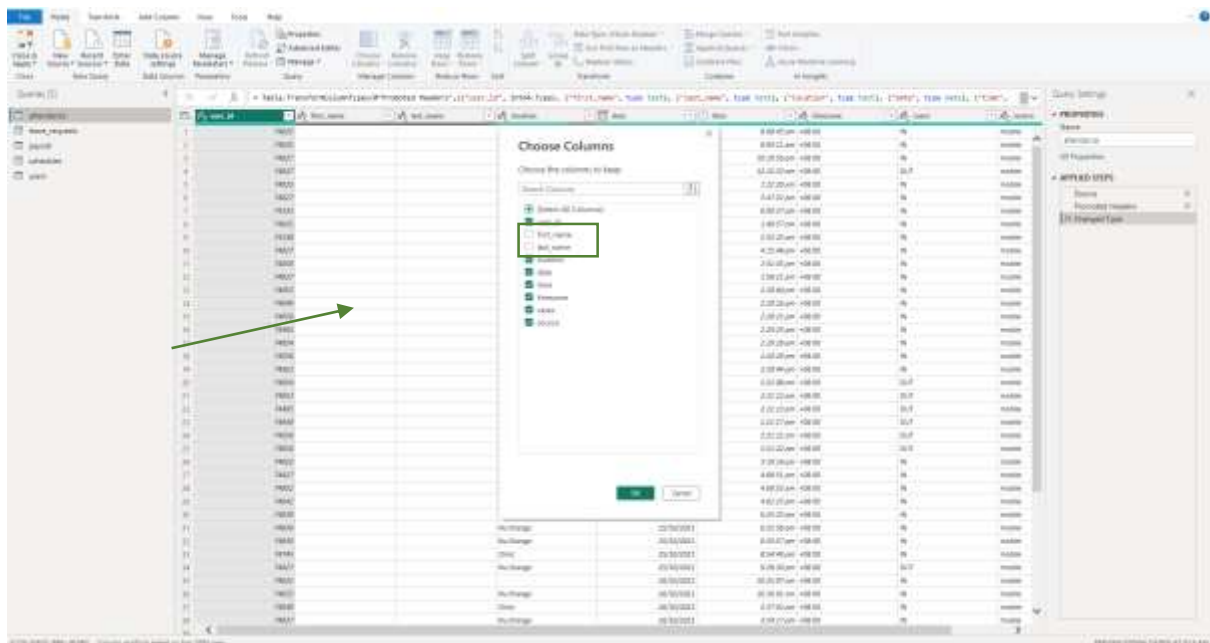


CLEANING AND TRANSFORMING ATTENDANCE TABLE

Removing unwanted columns attendance table. Select the attendance table in the Navigator Pane. Go to Home, click Choose Columns drop down arrow and click Choose Columns.



The Choose Columns pop up window appear. Unselect the check box that you want to remove from the attendance table. We chose to remove the first_name and last_name columns from the table because they have no data.



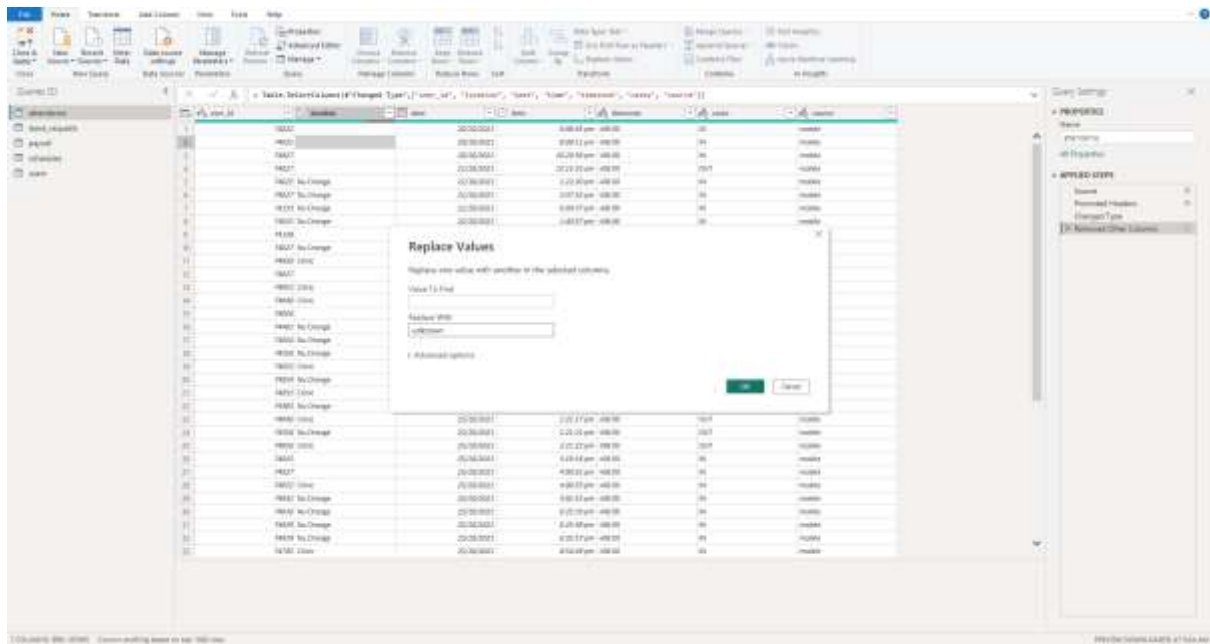
Removing unwanted columns result. See new applied steps.

The screenshot shows the QlikView interface with a data table. The table has columns: Location, Date, Price, Volume, and Status. The 'Location' column contains various codes like '10000', '10001', etc. The 'Status' column contains values like 'No Change', 'Change', etc. On the right, the 'Applied Steps' pane is visible, showing a list of steps. A green arrow points to the 'Remove Other Columns' step in the 'Applied Steps' pane.

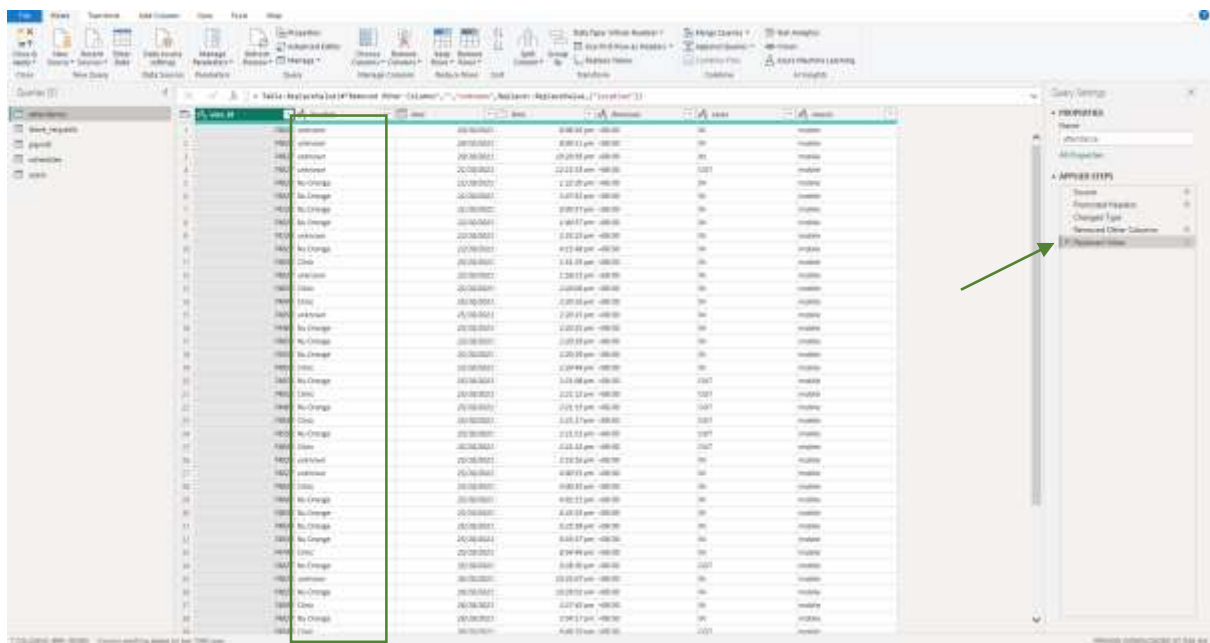
Replacing blank value of location column with "unknown". Right click the cell of location column and select replace value.

The screenshot shows the QlikView interface with the same data table. A right-click context menu is open over a cell in the 'Location' column. The menu options include 'Copy', 'Paste', 'Test Filter', 'Apply Filter', 'Clear Filter', 'Add as New Column', and 'Replace Value'. The 'Replace Value' option is highlighted.

Leave the Value to Find input box blank. Input unknown in Replace With and click ok.

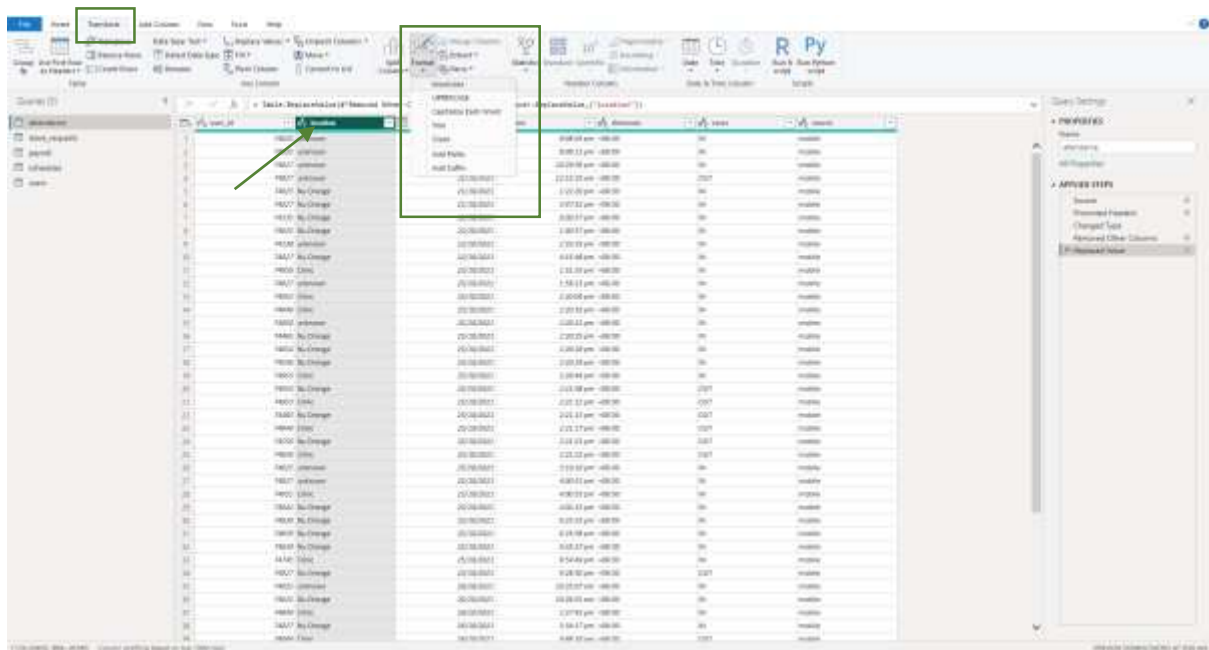


Result of replacing the blank value with unknown in location column.

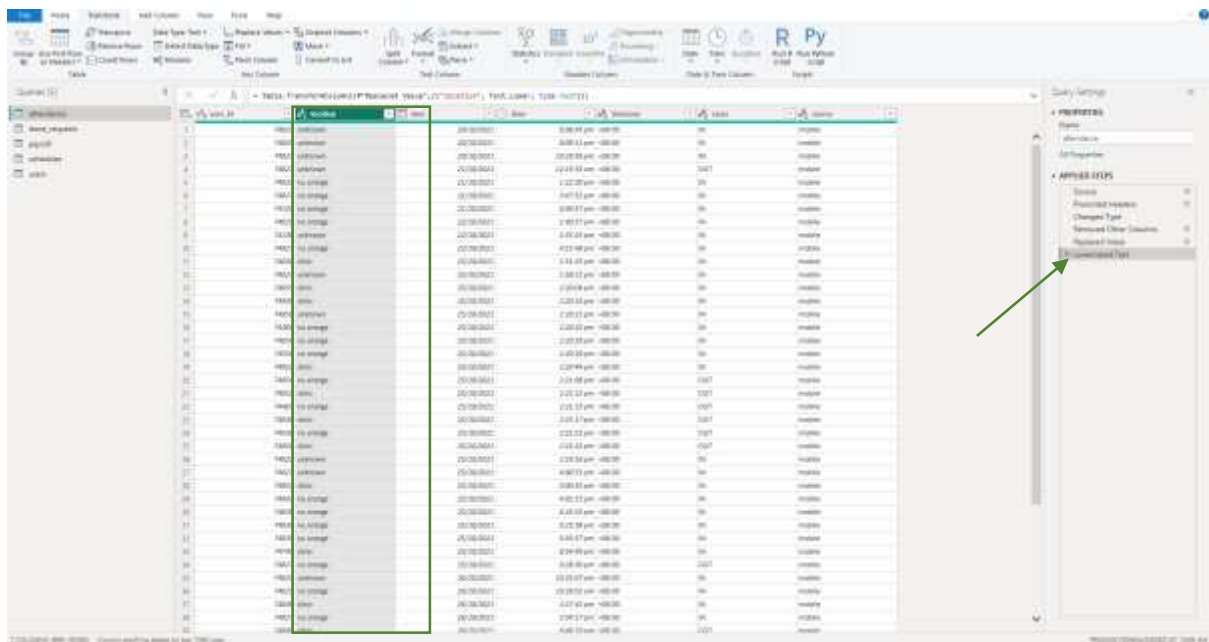


Changing the text format of the location column to lowercase letters.

Highlight the location column. Go to Transform, find and click the drop down arrow of Format and click lowercase.

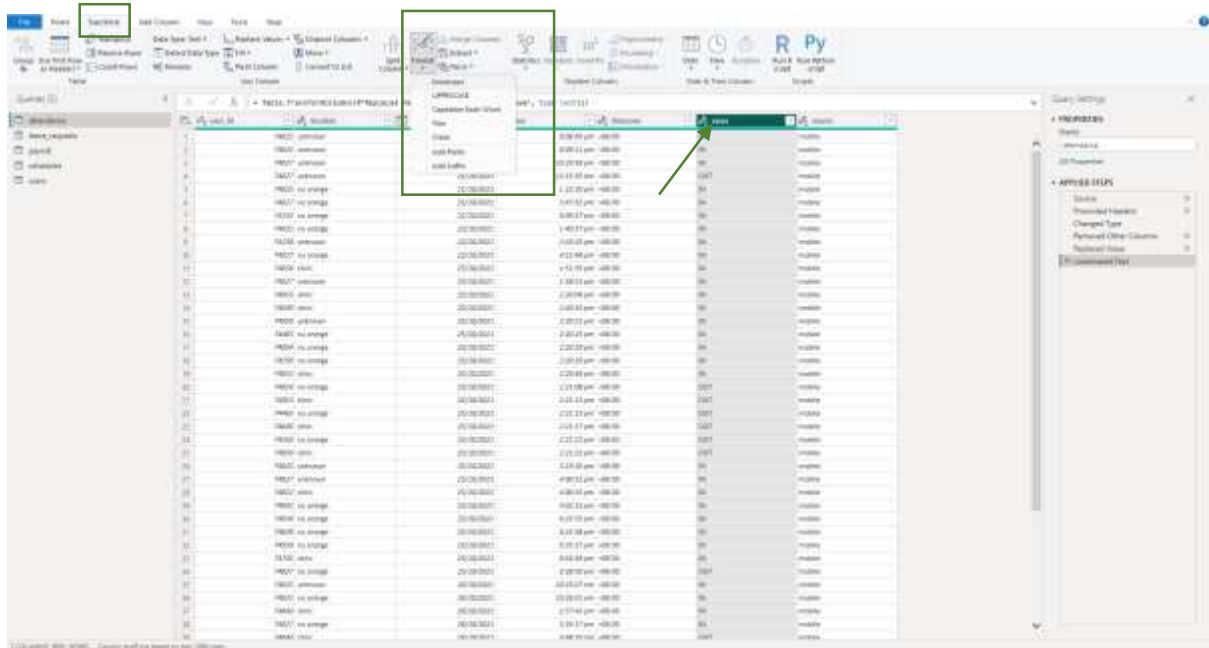


Result of location column formatted to lowercase letter.

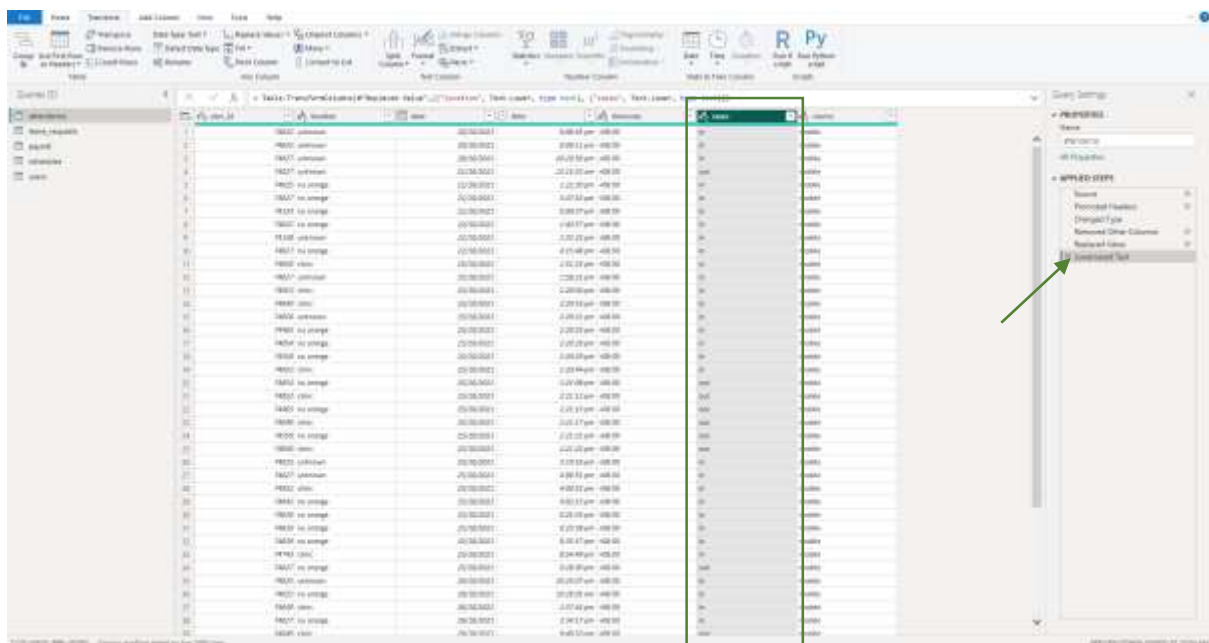


Changing the text format of the cases column to lowercase letters.

Highlight the cases column. Go to Transform, find and click the drop down arrow of Format and click lowercase.

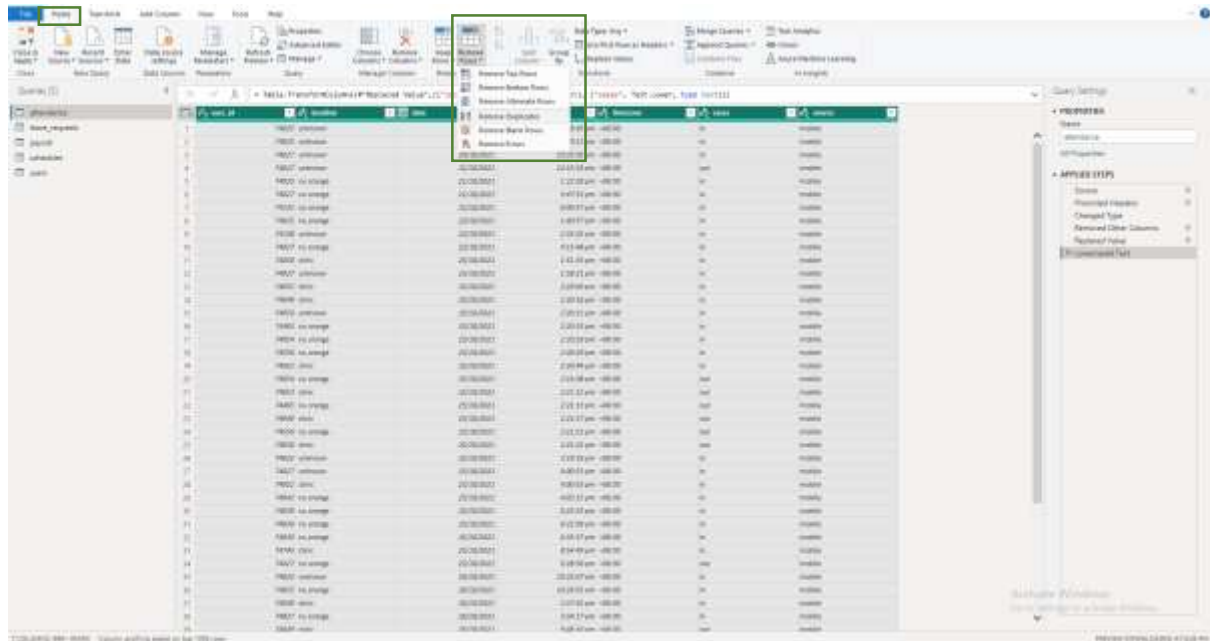


Result of cases column formatted to lowercase letters.

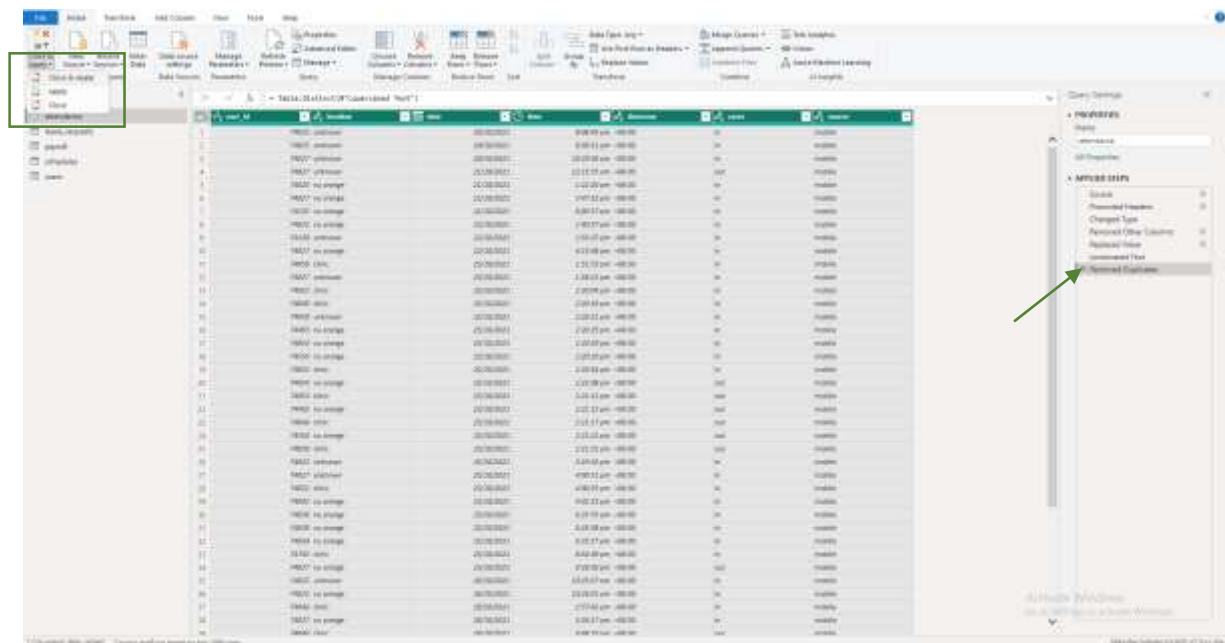


Removing duplicate of attendance table.

Highlight all the column of attendance table, go to Home ,click the Remove Rows drop down arrow and click Remove Duplicates.

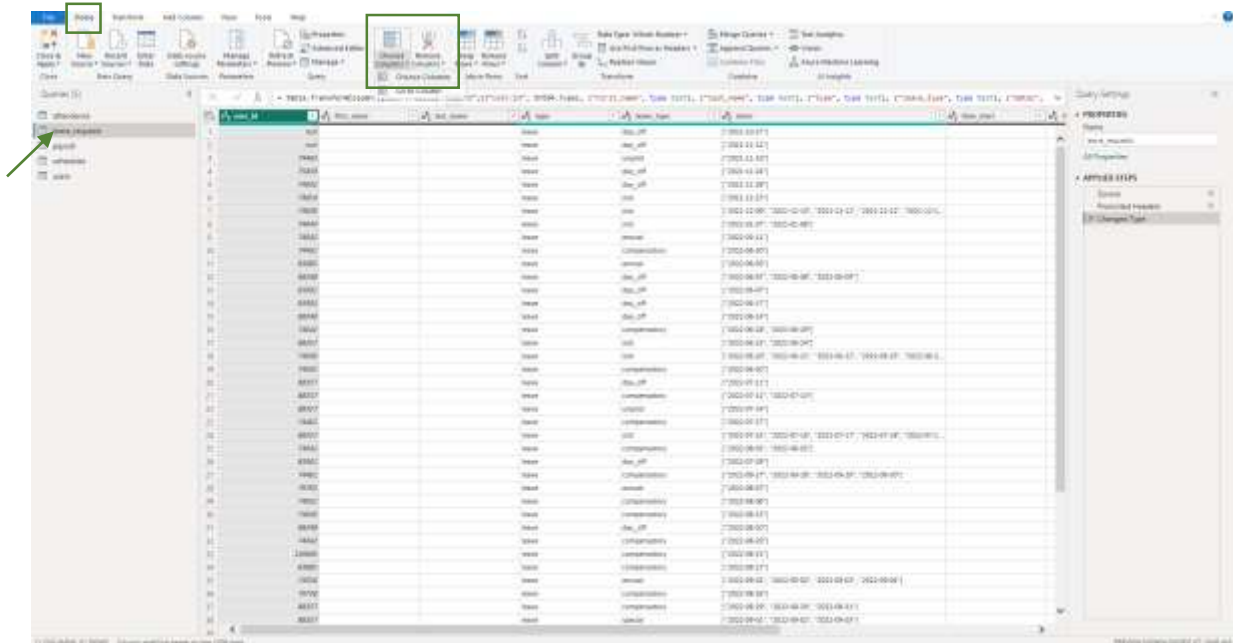


Result of Removing Duplicates of attendance table with 12604 rows. Click close and apply to load the data to model.

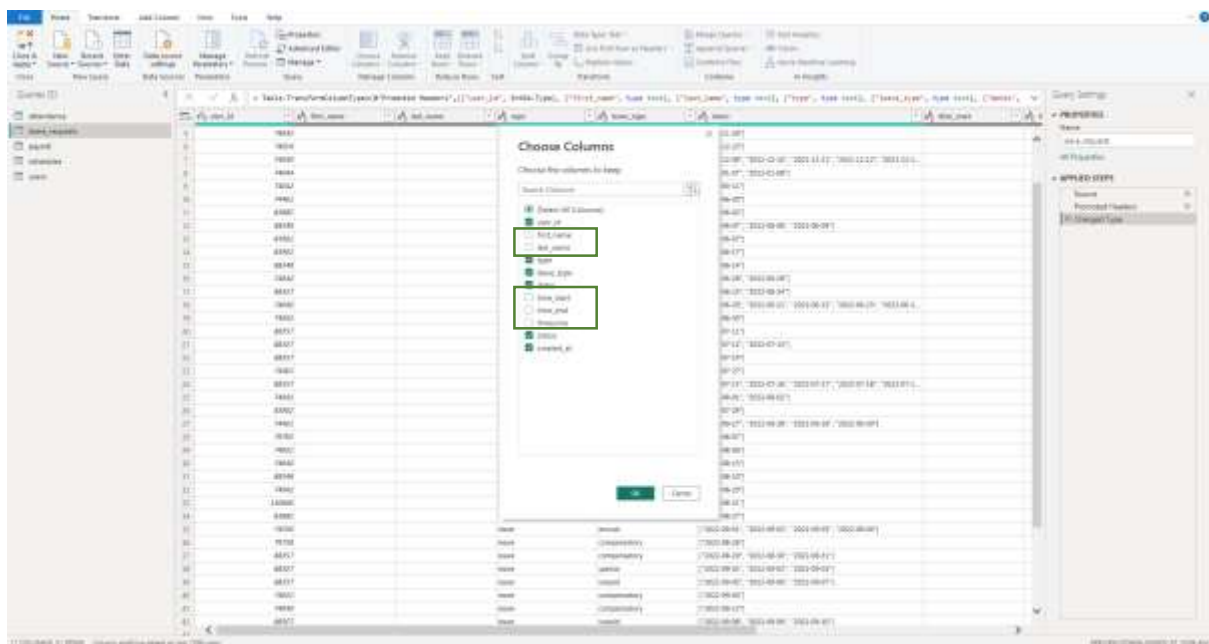


CLEANING AND TRANSFORMING LEAVE REQUEST TABLE

Removing unwanted columns of leave_requests table. Select the leave_requests table in the Navigator Pane. Go to Home, click Choose Columns dropdown arrow and then click Choose Columns.



The Choose Columns pop up window will appear. Unselect the check box that you want to remove from the leave_requests table. We chose to remove the first_name, last_name, time_start, time_end and timezone columns from the table because they have no data.



Result of removing unwanted column from leave_requests table.

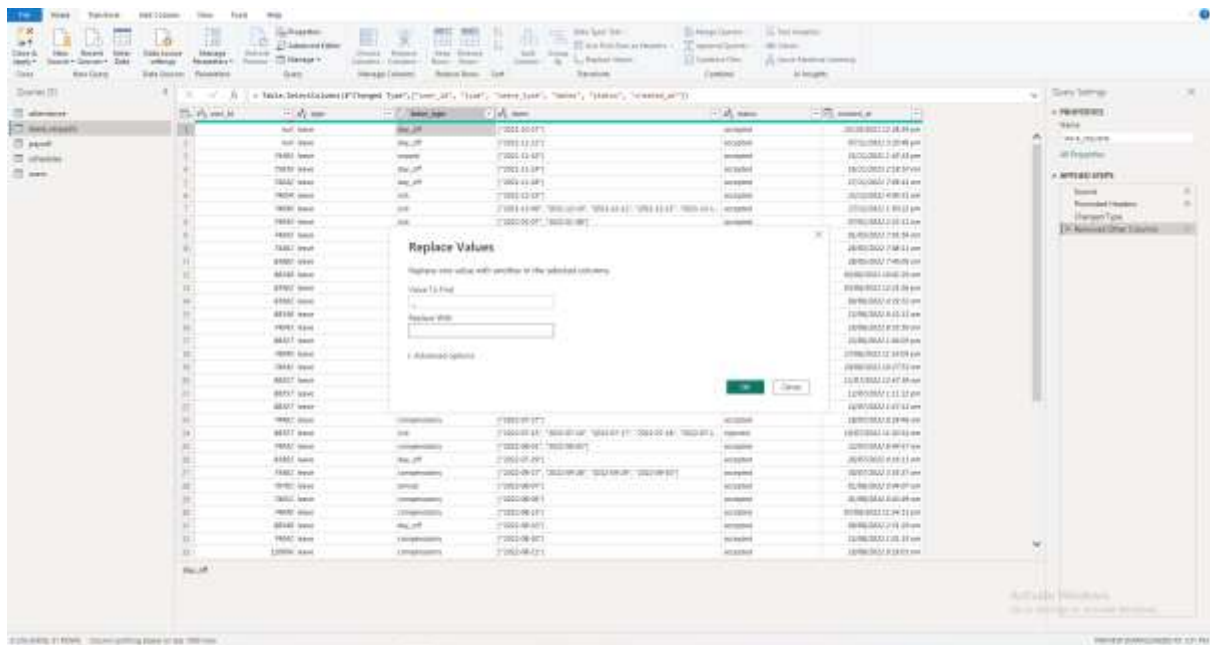
id	name	day_off	leave_type	amount
1	John Doe	2023-01-01	vacation	10
2	Jane Smith	2023-01-02	sick	5
3	Mike Johnson	2023-01-03	compensation	15
4	Sarah Brown	2023-01-04	vacation	8
5	David Wilson	2023-01-05	sick	3
6	Emily Davis	2023-01-06	compensation	12
7	Chris Miller	2023-01-07	vacation	7
8	Alice Taylor	2023-01-08	sick	4
9	Bob Anderson	2023-01-09	compensation	11
10	Charlie White	2023-01-10	vacation	9
11	Diana Black	2023-01-11	sick	6
12	Frank Green	2023-01-12	compensation	14
13	Grace Hall	2023-01-13	vacation	6
14	Henry King	2023-01-14	sick	2
15	Ivy Lee	2023-01-15	compensation	13
16	Jack Reed	2023-01-16	vacation	5
17	Karen Scott	2023-01-17	sick	7
18	Leo Turner	2023-01-18	compensation	10
19	Mia Young	2023-01-19	vacation	4
20	Noah Adams	2023-01-20	sick	1
21	Olivia Baker	2023-01-21	compensation	9
22	Peter Clark	2023-01-22	vacation	3
23	Quinn Evans	2023-01-23	sick	8
24	Rachel Fisher	2023-01-24	compensation	6
25	Samuel Hill	2023-01-25	vacation	2
26	Tina King	2023-01-26	sick	5
27	Uma Lee	2023-01-27	compensation	7
28	Victor Miller	2023-01-28	vacation	1
29	Wendy Moore	2023-01-29	sick	4
30	Xavier Nelson	2023-01-30	compensation	11
31	Yara Ortiz	2023-01-31	vacation	6
32	Zoe Parker	2023-02-01	sick	3
33	Adam Scott	2023-02-02	compensation	8
34	Bella Turner	2023-02-03	vacation	5
35	Carl White	2023-02-04	sick	2
36	Dora Young	2023-02-05	compensation	7
37	Ethan Adams	2023-02-06	vacation	4
38	Fiona Baker	2023-02-07	sick	1
39	Gavin Clark	2023-02-08	compensation	6
40	Hannah Evans	2023-02-09	vacation	3
41	Ian Fisher	2023-02-10	sick	8
42	Jessica Hill	2023-02-11	compensation	5
43	Kyle King	2023-02-12	vacation	2
44	Laura Lee	2023-02-13	sick	7
45	Mark Miller	2023-02-14	compensation	4
46	Nancy Moore	2023-02-15	vacation	1
47	Oscar Nelson	2023-02-16	sick	6
48	Pamela Ortiz	2023-02-17	compensation	3
49	Quinn Parker	2023-02-18	vacation	8
50	Rachel Scott	2023-02-19	sick	5
51	Samuel Turner	2023-02-20	compensation	2
52	Tina White	2023-02-21	vacation	7
53	Uma Young	2023-02-22	sick	4
54	Victor Adams	2023-02-23	compensation	1
55	Wendy Baker	2023-02-24	vacation	6
56	Xavier Clark	2023-02-25	sick	3
57	Yara Evans	2023-02-26	compensation	8
58	Zoe Fisher	2023-02-27	vacation	5
59	Adam Hill	2023-02-28	sick	2
60	Bella King	2023-02-29	compensation	7
61	Carl Lee	2023-03-01	vacation	4
62	Dora Miller	2023-03-02	sick	1
63	Ethan Moore	2023-03-03	compensation	6
64	Fiona Nelson	2023-03-04	vacation	3
65	Gavin Ortiz	2023-03-05	sick	8
66	Hannah Parker	2023-03-06	compensation	5
67	Ian Scott	2023-03-07	vacation	2
68	Jessica Turner	2023-03-08	sick	7
69	Kyle White	2023-03-09	compensation	4
70	Laura Young	2023-03-10	vacation	1
71	Mark Adams	2023-03-11	sick	6
72	Nancy Baker	2023-03-12	compensation	3
73	Oscar Clark	2023-03-13	vacation	8
74	Pamela Evans	2023-03-14	sick	5
75	Quinn Fisher	2023-03-15	compensation	2
76	Rachel Hill	2023-03-16	vacation	7
77	Samuel King	2023-03-17	sick	4
78	Tina Lee	2023-03-18	compensation	1
79	Uma Miller	2023-03-19	vacation	6
80	Victor Moore	2023-03-20	sick	3
81	Wendy Nelson	2023-03-21	compensation	8
82	Xavier Ortiz	2023-03-22	vacation	5
83	Yara Parker	2023-03-23	sick	2
84	Zoe Scott	2023-03-24	compensation	7
85	Adam Turner	2023-03-25	vacation	4
86	Bella White	2023-03-26	sick	1
87	Carl Young	2023-03-27	compensation	6
88	Dora Adams	2023-03-28	vacation	3
89	Ethan Baker	2023-03-29	sick	8
90	Fiona Clark	2023-03-30	compensation	5
91	Gavin Evans	2023-03-31	vacation	2
92	Hannah Fisher	2023-04-01	sick	7
93	Ian Hill	2023-04-02	compensation	4
94	Jessica King	2023-04-03	vacation	1
95	Kyle Lee	2023-04-04	sick	6
96	Laura Miller	2023-04-05	compensation	3
97	Mark Moore	2023-04-06	vacation	8
98	Nancy Nelson	2023-04-07	sick	5
99	Oscar Ortiz	2023-04-08	compensation	2
100	Pamela Parker	2023-04-09	vacation	7

Replace the underscore of day_off with space in leave_type.

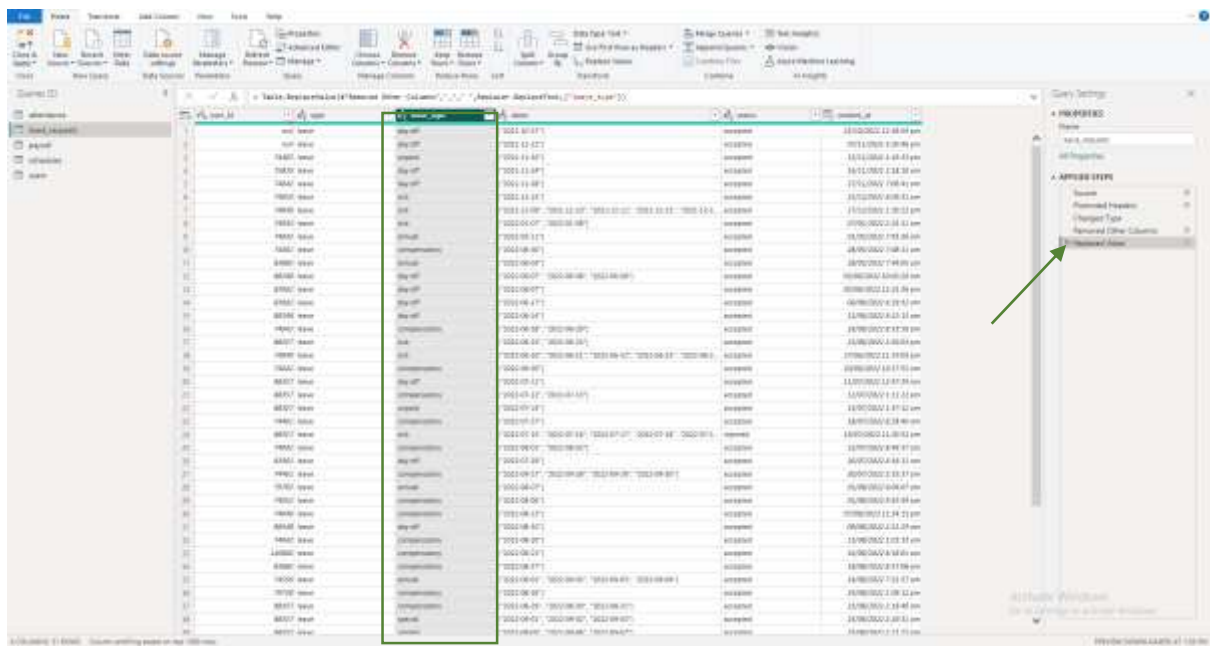
Right Click the cell of leave_type column which has a value of day_off, right click and select replace value.

id	name	day_off	leave_type	amount
1	John Doe	2023-01-01	vacation	10
2	Jane Smith	2023-01-02	sick	5
3	Mike Johnson	2023-01-03	compensation	15
4	Sarah Brown	2023-01-04	vacation	8
5	David Wilson	2023-01-05	sick	3
6	Emily Davis	2023-01-06	compensation	12
7	Chris Miller	2023-01-07	vacation	7
8	Alice Taylor	2023-01-08	sick	4
9	Bob Anderson	2023-01-09	compensation	11
10	Charlie White	2023-01-10	vacation	9
11	Diana Black	2023-01-11	sick	6
12	Frank Green	2023-01-12	compensation	14
13	Grace Hall	2023-01-13	vacation	6
14	Henry King	2023-01-14	sick	2
15	Ivy Lee	2023-01-15	compensation	13
16	Jack Reed	2023-01-16	vacation	5
17	Karen Scott	2023-01-17	sick	7
18	Leo Turner	2023-01-18	compensation	10
19	Mia Young	2023-01-19	vacation	4
20	Noah Adams	2023-01-20	sick	1
21	Olivia Baker	2023-01-21	compensation	9
22	Peter Clark	2023-01-22	vacation	3
23	Quinn Evans	2023-01-23	sick	8
24	Rachel Fisher	2023-01-24	compensation	6
25	Samuel Hill	2023-01-25	vacation	2
26	Tina King	2023-01-26	sick	5
27	Uma Lee	2023-01-27	compensation	7
28	Victor Miller	2023-01-28	vacation	1
29	Wendy Moore	2023-01-29	sick	4
30	Xavier Nelson	2023-01-30	compensation	11
31	Yara Ortiz	2023-01-31	vacation	6
32	Zoe Parker	2023-02-01	sick	3
33	Adam Scott	2023-02-02	compensation	8
34	Bella Turner	2023-02-03	vacation	5
35	Carl White	2023-02-04	sick	2
36	Dora Young	2023-02-05	compensation	7
37	Ethan Adams	2023-02-06	vacation	4
38	Fiona Baker	2023-02-07	sick	1
39	Gavin Clark	2023-02-08	compensation	6
40	Hannah Evans	2023-02-09	vacation	3
41	Ian Fisher	2023-02-10	sick	8
42	Jessica Hill	2023-02-11	compensation	5
43	Kyle King	2023-02-12	vacation	2
44	Laura Lee	2023-02-13	sick	7
45	Mark Miller	2023-02-14	compensation	4
46	Nancy Moore	2023-02-15	vacation	1
47	Oscar Nelson	2023-02-16	sick	6
48	Pamela Ortiz	2023-02-17	compensation	3
49	Quinn Parker	2023-02-18	vacation	8
50	Rachel Scott	2023-02-19	sick	5
51	Samuel Turner	2023-02-20	compensation	2
52	Tina White	2023-02-21	vacation	7
53	Uma Young	2023-02-22	sick	4
54	Victor Adams	2023-02-23	compensation	1
55	Wendy Baker	2023-02-24	vacation	6
56	Xavier Clark	2023-02-25	sick	3
57	Yara Evans	2023-02-26	compensation	8
58	Zoe Fisher	2023-02-27	vacation	5
59	Adam Hill	2023-02-28	sick	2
60	Bella King	2023-02-29	compensation	7
61	Carl Lee	2023-03-01	vacation	4
62	Dora Miller	2023-03-02	sick	1
63	Ethan Moore	2023-03-03	compensation	6
64	Fiona Nelson	2023-03-04	vacation	3
65	Gavin Ortiz	2023-03-05	sick	8
66	Hannah Parker	2023-03-06	compensation	5
67	Ian Scott	2023-03-07	vacation	2
68	Jessica Turner	2023-03-08	sick	7
69	Kyle White	2023-03-09	compensation	4
70	Laura Young	2023-03-10	vacation	1
71	Mark Adams	2023-03-11	sick	6
72	Nancy Baker	2023-03-12	compensation	3
73	Oscar Clark	2023-03-13	vacation	8
74	Pamela Evans	2023-03-14	sick	5
75	Quinn Fisher	2023-03-15	compensation	2
76	Rachel Hill	2023-03-16	vacation	7
77	Samuel King	2023-03-17	sick	4
78	Tina Lee	2023-03-18	compensation	1
79	Uma Miller	2023-03-19	vacation	6
80	Victor Moore	2023-03-20	sick	3
81	Wendy Nelson	2023-03-21	compensation	8
82	Xavier Ortiz	2023-03-22	vacation	5
83	Yara Parker	2023-03-23	sick	2
84	Zoe Scott	2023-03-24	compensation	7
85	Adam Turner	2023-03-25	vacation	4
86	Bella White	2023-03-26	sick	1
87	Carl Young	2023-03-27	compensation	6
88	Dora Adams	2023-03-28	vacation	3
89	Ethan Baker	2023-03-29	sick	8
90	Fiona Clark	2023-03-30	compensation	5
91	Gavin Evans	2023-03-31	vacation	2
92	Hannah Fisher	2023-04-01	sick	7
93	Ian Hill	2023-04-02	compensation	4
94	Jessica King	2023-04-03	vacation	1
95	Kyle Lee	2023-04-04	sick	6
96	Laura Miller	2023-04-05	compensation	3
97	Mark Moore	2023-04-06	vacation	8
98	Nancy Nelson	2023-04-07	sick	5
99	Oscar Ortiz	2023-04-08	compensation	2
100	Pamela Parker	2023-04-09	vacation	7

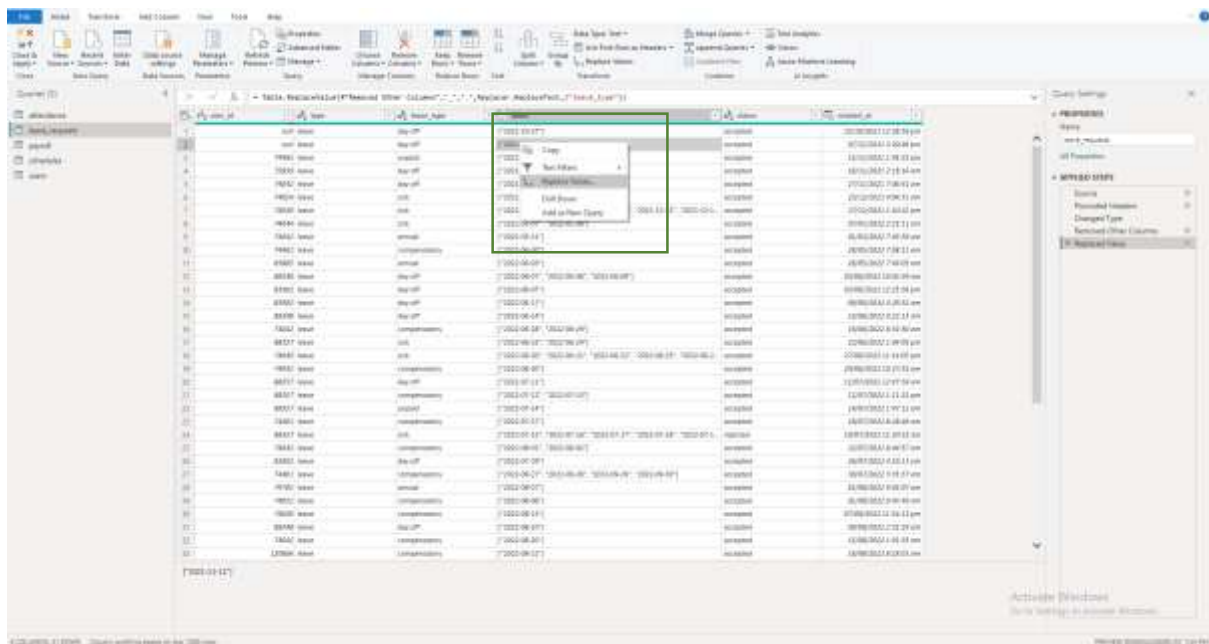
Input “_” to Value to Find and inside the Replace With input box press the space bar and click ok.



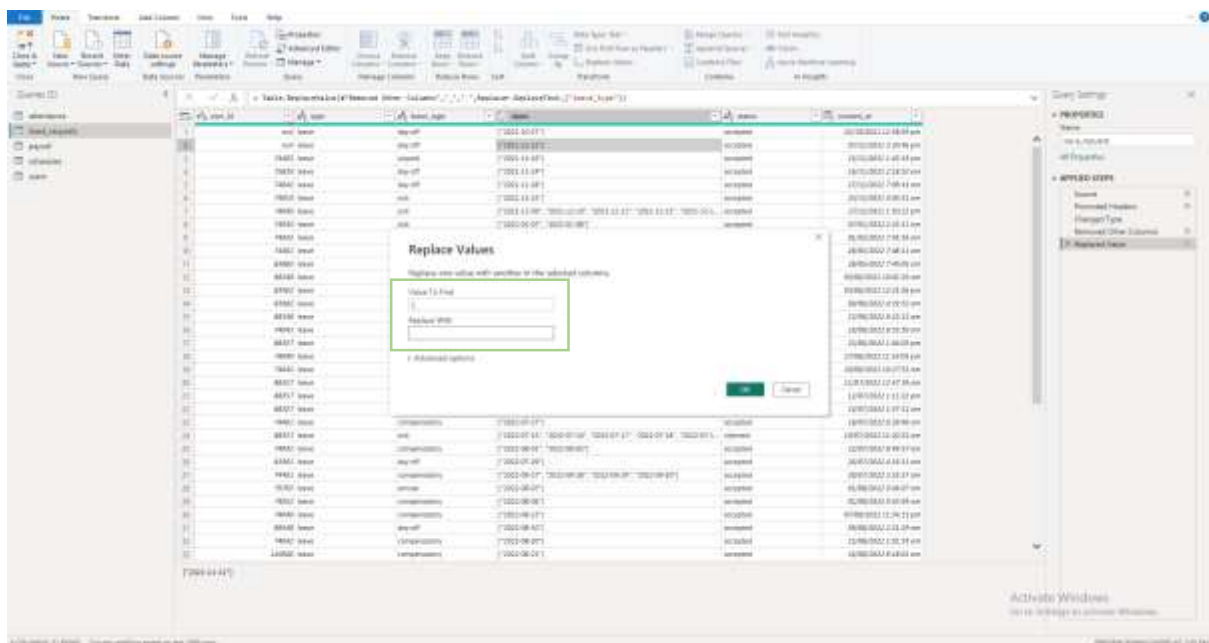
Result of replacing the day off underscore by space.



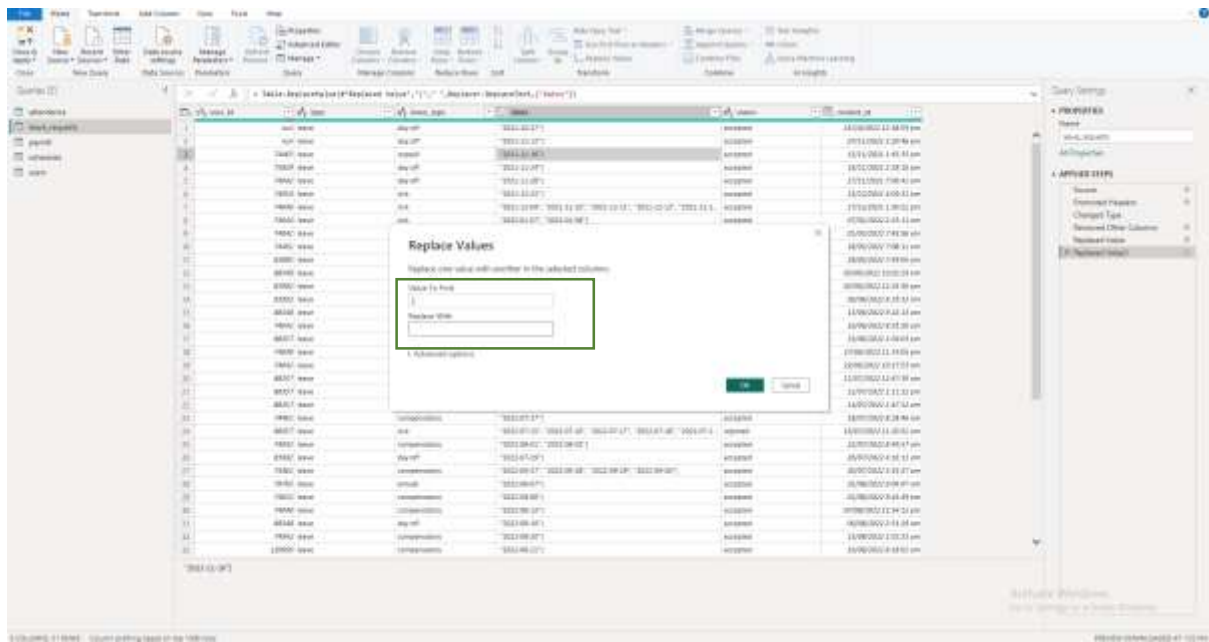
Cleaning the date column of the leave_requests table. Right click the cell of the dates column and click replace values.



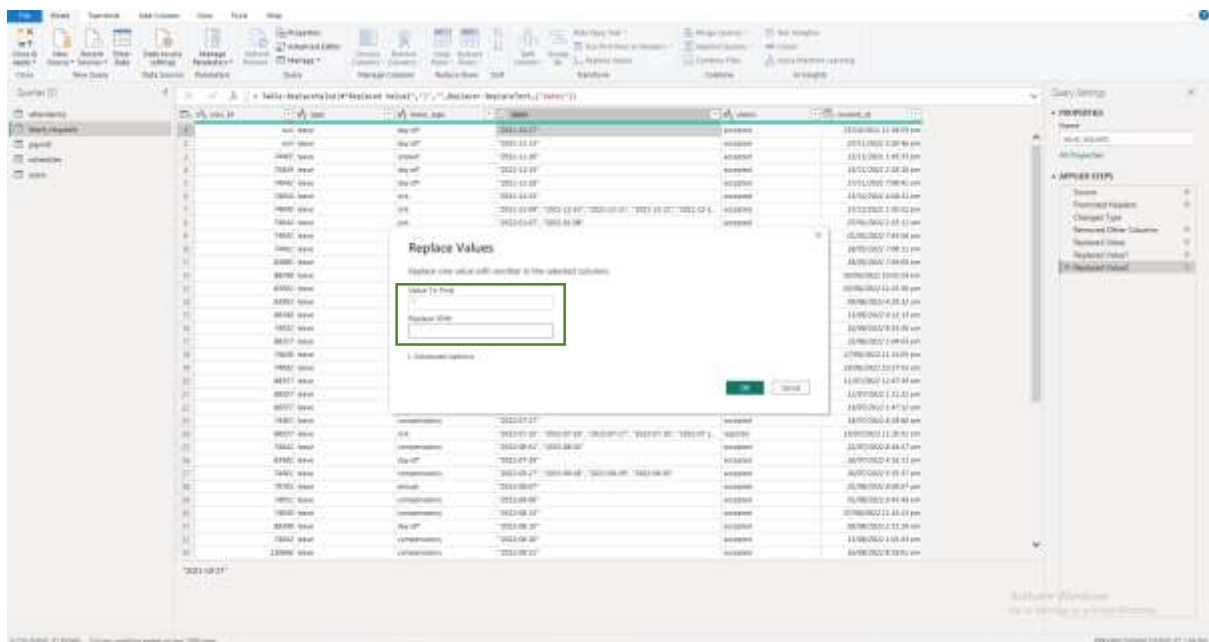
Input the open bracket to Values to Find and leave the Replace With input box blank and then click ok.



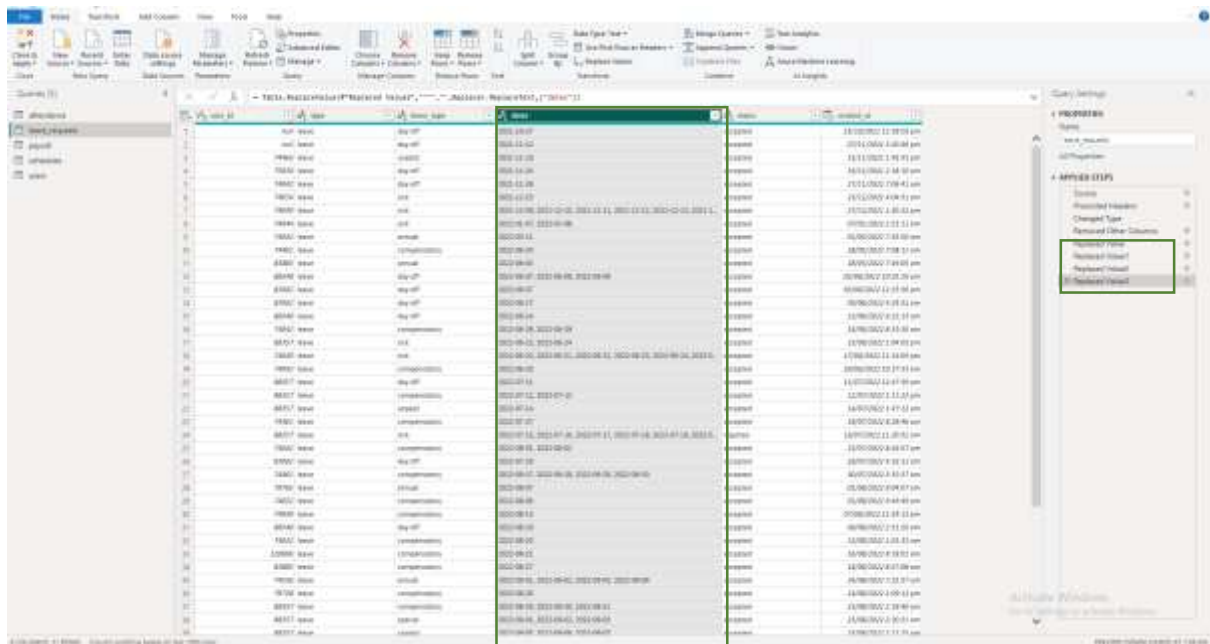
Right click again the cell of the dates column and select replace values. Input the close bracket to Values to Find and leave the Replace With input box blank and then click ok.



Right click again the cell of the dates column and select replace values. Input the double quote to Values to Find and leave the Replace With input box blank and then click ok.

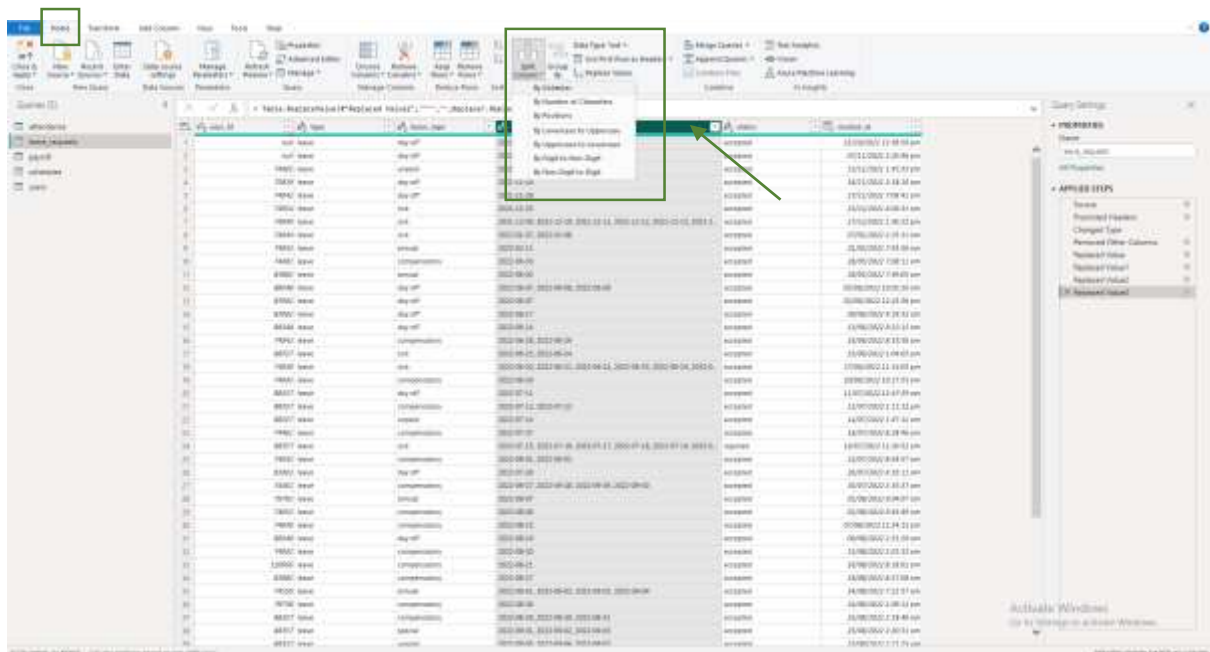


This is the result of removing the open and close brackets and double quotes.



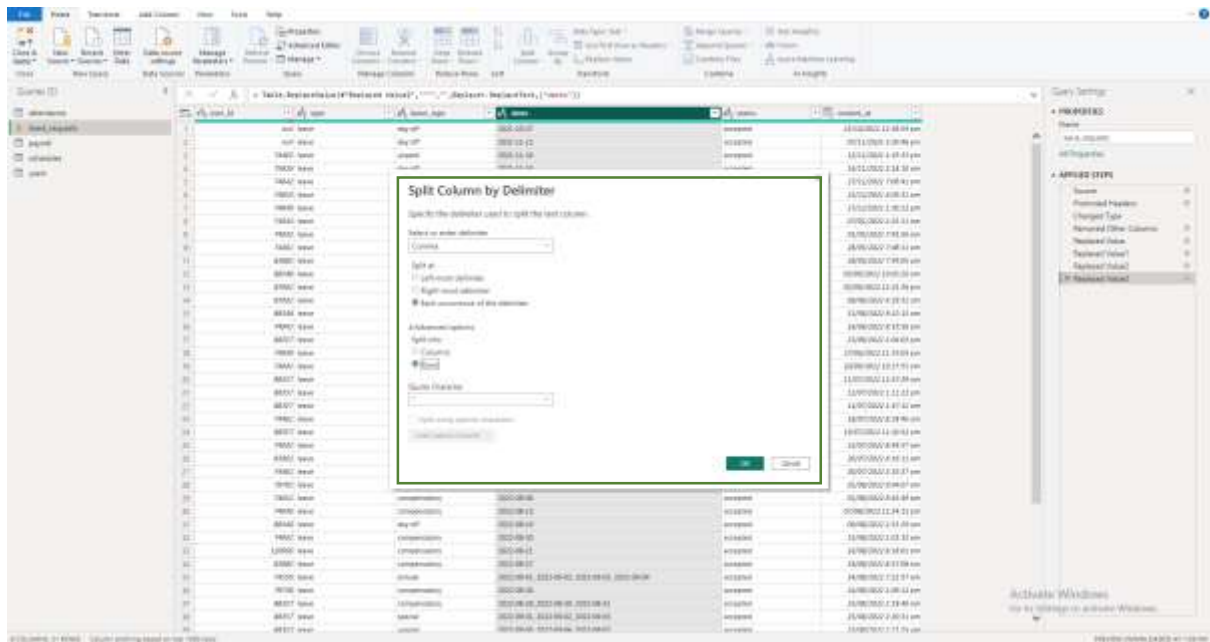
Name	Status	Date	Address
1001	open	2012-01-01	1001-01-01
1002	open	2012-01-02	1002-01-02
1003	open	2012-01-03	1003-01-03
1004	open	2012-01-04	1004-01-04
1005	open	2012-01-05	1005-01-05
1006	open	2012-01-06	1006-01-06
1007	open	2012-01-07	1007-01-07
1008	open	2012-01-08	1008-01-08
1009	open	2012-01-09	1009-01-09
1010	open	2012-01-10	1010-01-10
1011	open	2012-01-11	1011-01-11
1012	open	2012-01-12	1012-01-12
1013	open	2012-01-13	1013-01-13
1014	open	2012-01-14	1014-01-14
1015	open	2012-01-15	1015-01-15
1016	open	2012-01-16	1016-01-16
1017	open	2012-01-17	1017-01-17
1018	open	2012-01-18	1018-01-18
1019	open	2012-01-19	1019-01-19
1020	open	2012-01-20	1020-01-20
1021	open	2012-01-21	1021-01-21
1022	open	2012-01-22	1022-01-22
1023	open	2012-01-23	1023-01-23
1024	open	2012-01-24	1024-01-24
1025	open	2012-01-25	1025-01-25
1026	open	2012-01-26	1026-01-26
1027	open	2012-01-27	1027-01-27
1028	open	2012-01-28	1028-01-28
1029	open	2012-01-29	1029-01-29
1030	open	2012-01-30	1030-01-30
1031	open	2012-01-31	1031-01-31
1032	open	2012-02-01	1032-02-01
1033	open	2012-02-02	1033-02-02
1034	open	2012-02-03	1034-02-03
1035	open	2012-02-04	1035-02-04
1036	open	2012-02-05	1036-02-05
1037	open	2012-02-06	1037-02-06
1038	open	2012-02-07	1038-02-07
1039	open	2012-02-08	1039-02-08
1040	open	2012-02-09	1040-02-09
1041	open	2012-02-10	1041-02-10
1042	open	2012-02-11	1042-02-11
1043	open	2012-02-12	1043-02-12
1044	open	2012-02-13	1044-02-13
1045	open	2012-02-14	1045-02-14
1046	open	2012-02-15	1046-02-15
1047	open	2012-02-16	1047-02-16
1048	open	2012-02-17	1048-02-17
1049	open	2012-02-18	1049-02-18
1050	open	2012-02-19	1050-02-19
1051	open	2012-02-20	1051-02-20
1052	open	2012-02-21	1052-02-21
1053	open	2012-02-22	1053-02-22
1054	open	2012-02-23	1054-02-23
1055	open	2012-02-24	1055-02-24
1056	open	2012-02-25	1056-02-25
1057	open	2012-02-26	1057-02-26
1058	open	2012-02-27	1058-02-27
1059	open	2012-02-28	1059-02-28
1060	open	2012-02-29	1060-02-29

Splitting the date column by delimiter. Click to highlight the dates column, go to Home and click the Split Column dropdown arrow and select by Delimiter.

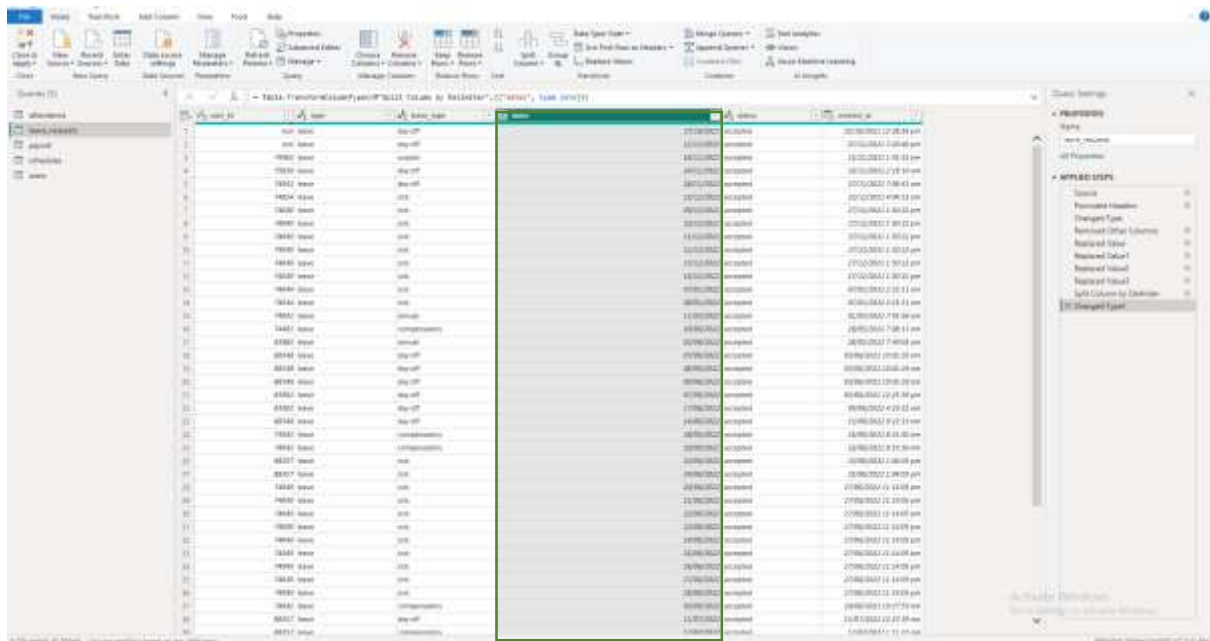


Name	Status	Date	Address
1001	open	2012-01-01	1001-01-01
1002	open	2012-01-02	1002-01-02
1003	open	2012-01-03	1003-01-03
1004	open	2012-01-04	1004-01-04
1005	open	2012-01-05	1005-01-05
1006	open	2012-01-06	1006-01-06
1007	open	2012-01-07	1007-01-07
1008	open	2012-01-08	1008-01-08
1009	open	2012-01-09	1009-01-09
1010	open	2012-01-10	1010-01-10
1011	open	2012-01-11	1011-01-11
1012	open	2012-01-12	1012-01-12
1013	open	2012-01-13	1013-01-13
1014	open	2012-01-14	1014-01-14
1015	open	2012-01-15	1015-01-15
1016	open	2012-01-16	1016-01-16
1017	open	2012-01-17	1017-01-17
1018	open	2012-01-18	1018-01-18
1019	open	2012-01-19	1019-01-19
1020	open	2012-01-20	1020-01-20
1021	open	2012-01-21	1021-01-21
1022	open	2012-01-22	1022-01-22
1023	open	2012-01-23	1023-01-23
1024	open	2012-01-24	1024-01-24
1025	open	2012-01-25	1025-01-25
1026	open	2012-01-26	1026-01-26
1027	open	2012-01-27	1027-01-27
1028	open	2012-01-28	1028-01-28
1029	open	2012-01-29	1029-01-29
1030	open	2012-01-30	1030-01-30
1031	open	2012-01-31	1031-01-31
1032	open	2012-02-01	1032-02-01
1033	open	2012-02-02	1033-02-02
1034	open	2012-02-03	1034-02-03
1035	open	2012-02-04	1035-02-04
1036	open	2012-02-05	1036-02-05
1037	open	2012-02-06	1037-02-06
1038	open	2012-02-07	1038-02-07
1039	open	2012-02-08	1039-02-08
1040	open	2012-02-09	1040-02-09
1041	open	2012-02-10	1041-02-10
1042	open	2012-02-11	1042-02-11
1043	open	2012-02-12	1043-02-12
1044	open	2012-02-13	1044-02-13
1045	open	2012-02-14	1045-02-14
1046	open	2012-02-15	1046-02-15
1047	open	2012-02-16	1047-02-16
1048	open	2012-02-17	1048-02-17
1049	open	2012-02-18	1049-02-18
1050	open	2012-02-19	1050-02-19
1051	open	2012-02-20	1051-02-20
1052	open	2012-02-21	1052-02-21
1053	open	2012-02-22	1053-02-22
1054	open	2012-02-23	1054-02-23
1055	open	2012-02-24	1055-02-24
1056	open	2012-02-25	1056-02-25
1057	open	2012-02-26	1057-02-26
1058	open	2012-02-27	1058-02-27
1059	open	2012-02-28	1059-02-28
1060	open	2012-02-29	1060-02-29

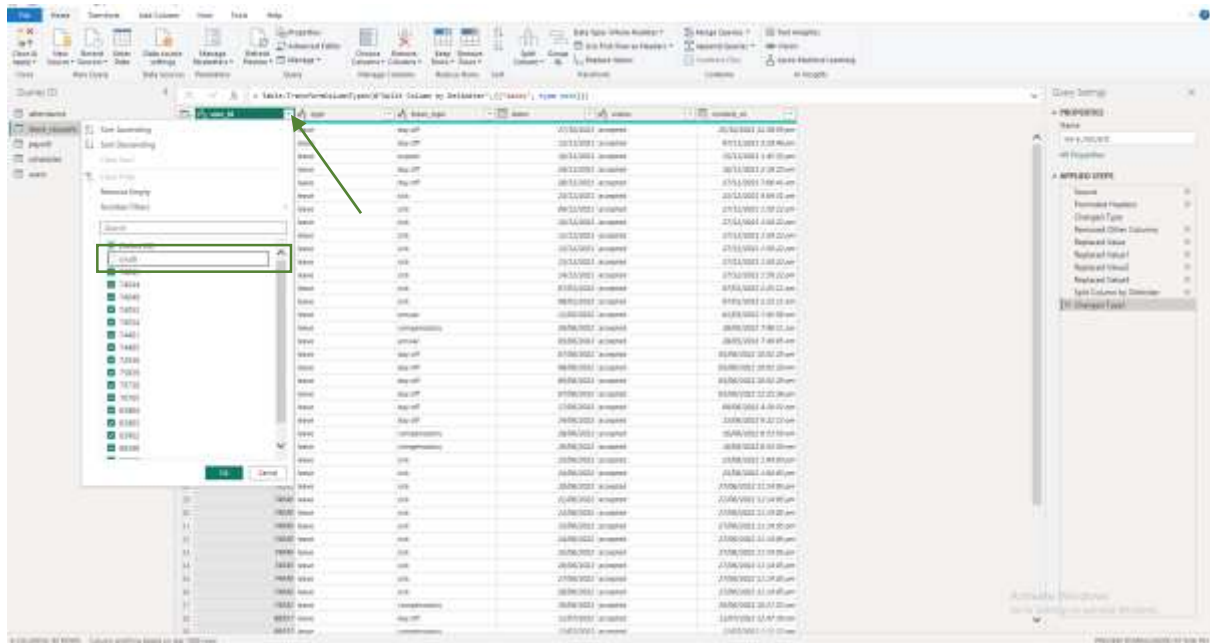
In the Split Column by delimiter pop up window, drop and select the “Comma” in the Select or enter delimiter. In the Split at choose “Each occurrence of the delimiter”. Click the Advance Options and in the split into choose “Rows” and then click ok.



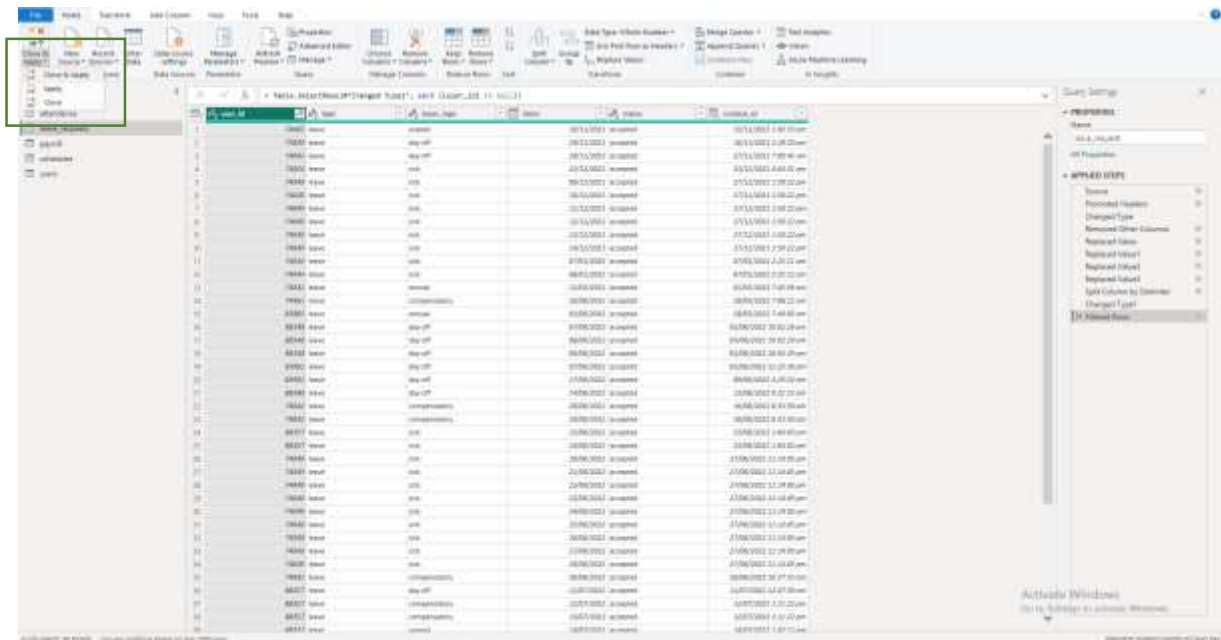
The result of Splitting the values of dates column by delimiter.



Removing the null value from the user_id column of leave_request table. Click the dropdown arrow from the user_id column and uncheck the null to filter it out of the table, and then click ok.

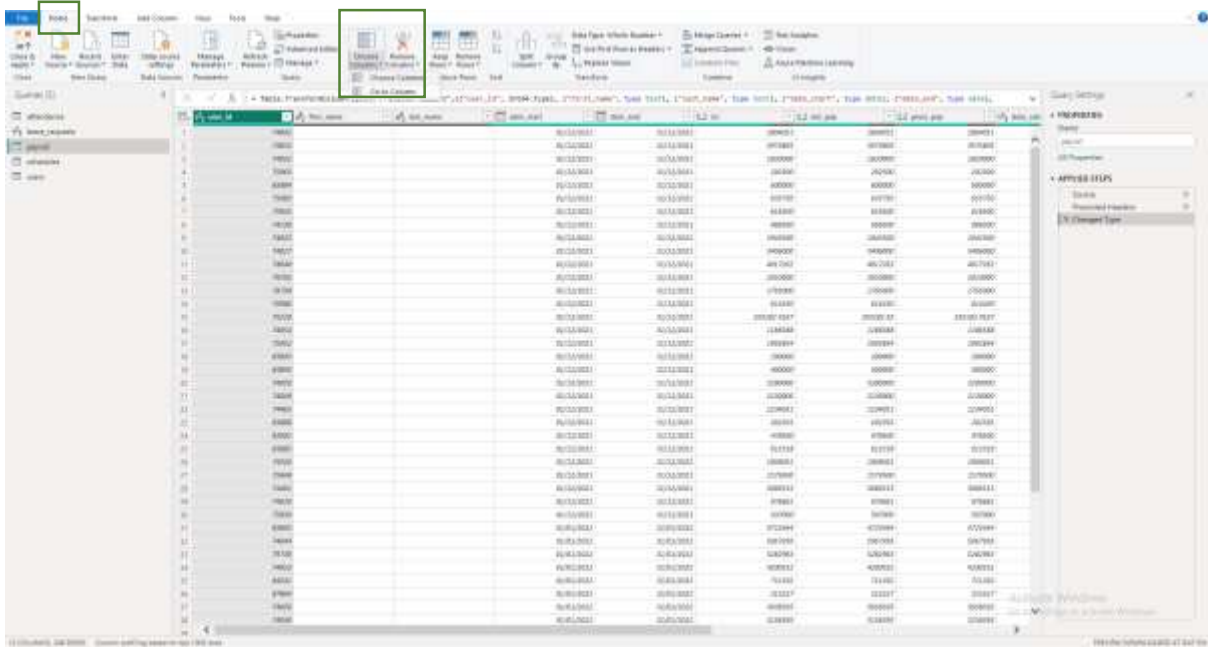


Go to and click Close and apply to load the data. The row count of leave_requests table is 90.

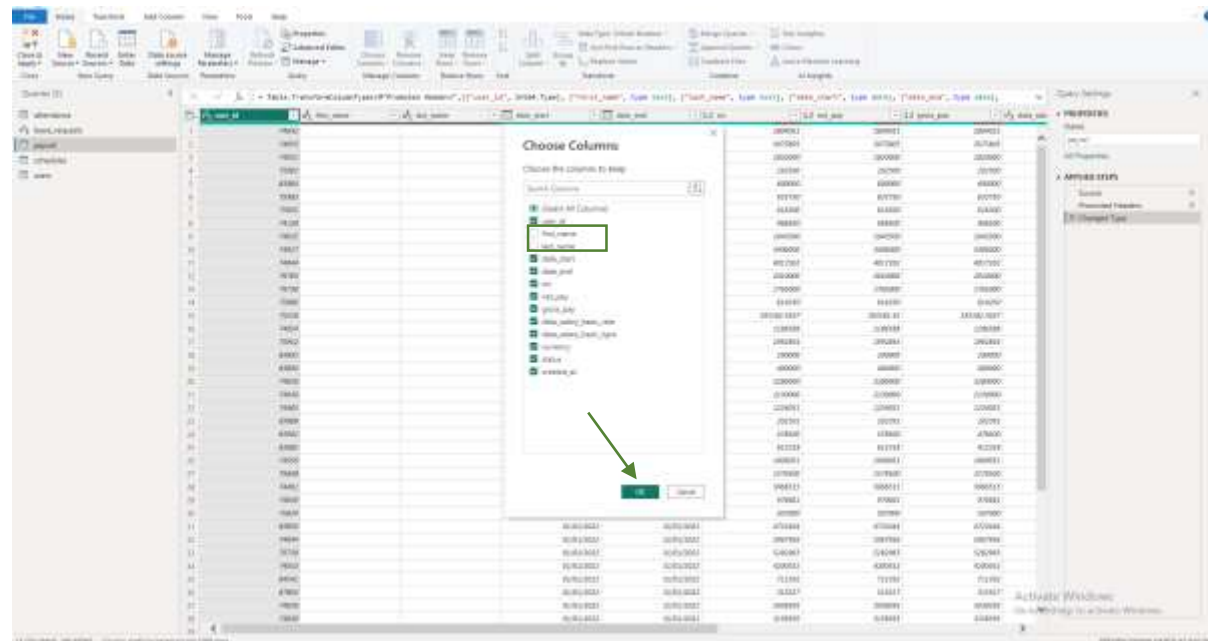


CLEANING AND TRANSFORMING PAYROLL TABLE

Remove unwanted columns from the payroll table. Go to Home and click the drop down arrow of Choose Columns and then select Choose Columns.



Unselect the check box of the column of first_name and last_name and then click ok.



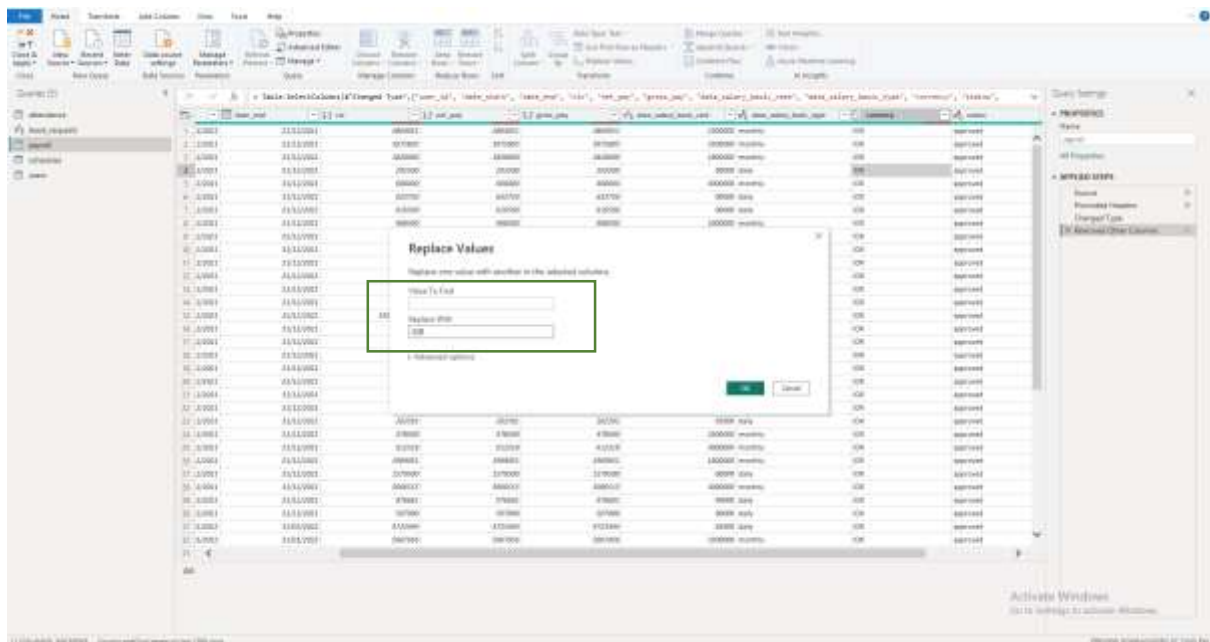
Result of removing columns first_name and last_name of payroll table.

The screenshot shows a database management tool interface. The main window displays a table with 20 rows and 14 columns. The columns are: emp_id, emp_name, emp_salary, emp_hire_date, emp_job_id, emp_dept_id, emp_status, emp_grade, emp_payroll_status, emp_payroll_type, emp_payroll_frequency, emp_payroll_start_date, emp_payroll_end_date, emp_payroll_amount, and emp_payroll_currency. The table contains data for various employees, including their names, salaries, hire dates, job IDs, department IDs, status, grade, payroll status, type, frequency, start and end dates, amount, and currency. A green arrow points to the 'emp_payroll_currency' column in the right-hand pane.

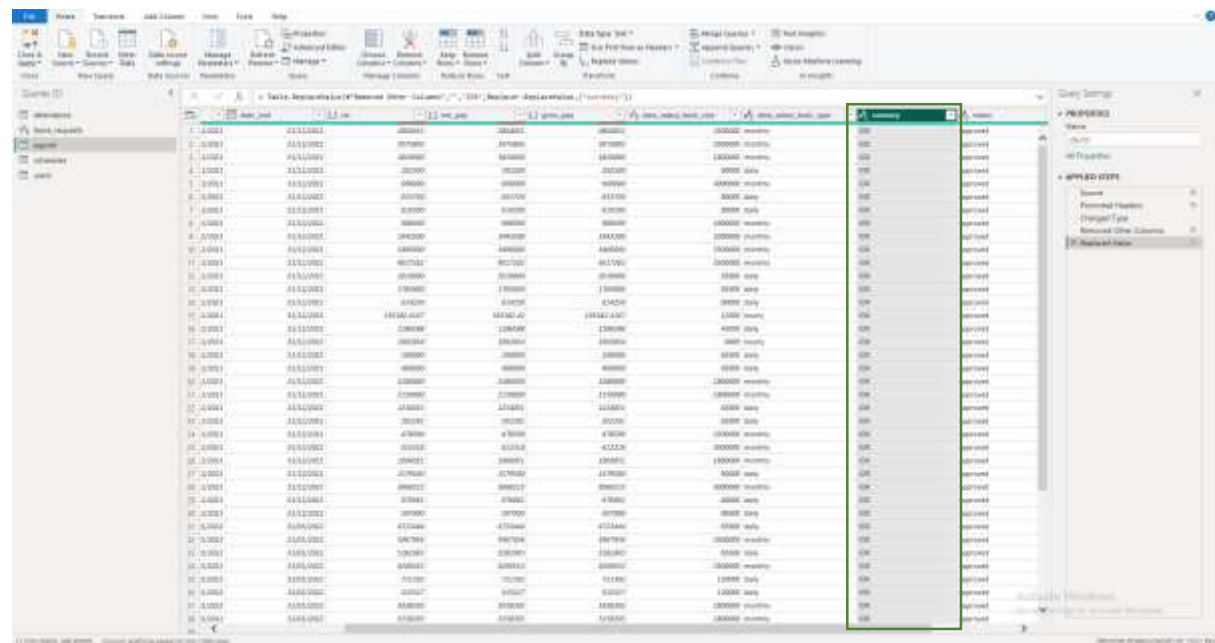
Populating the blank value of currency column with the IDR. Right Click the cell of the currency column and select the replace values.

The screenshot shows the same database management tool interface as above, but with a context menu open over the 'emp_payroll_currency' column. The context menu options are: Copy, Paste, Undo, Redo, Delete, Insert, Replace Values, and Set Default Value. The 'Replace Values' option is highlighted. The table data remains the same as in the previous screenshot.

Leave the Value to Find blank, put IDR to the input box of Replace With and then click ok.

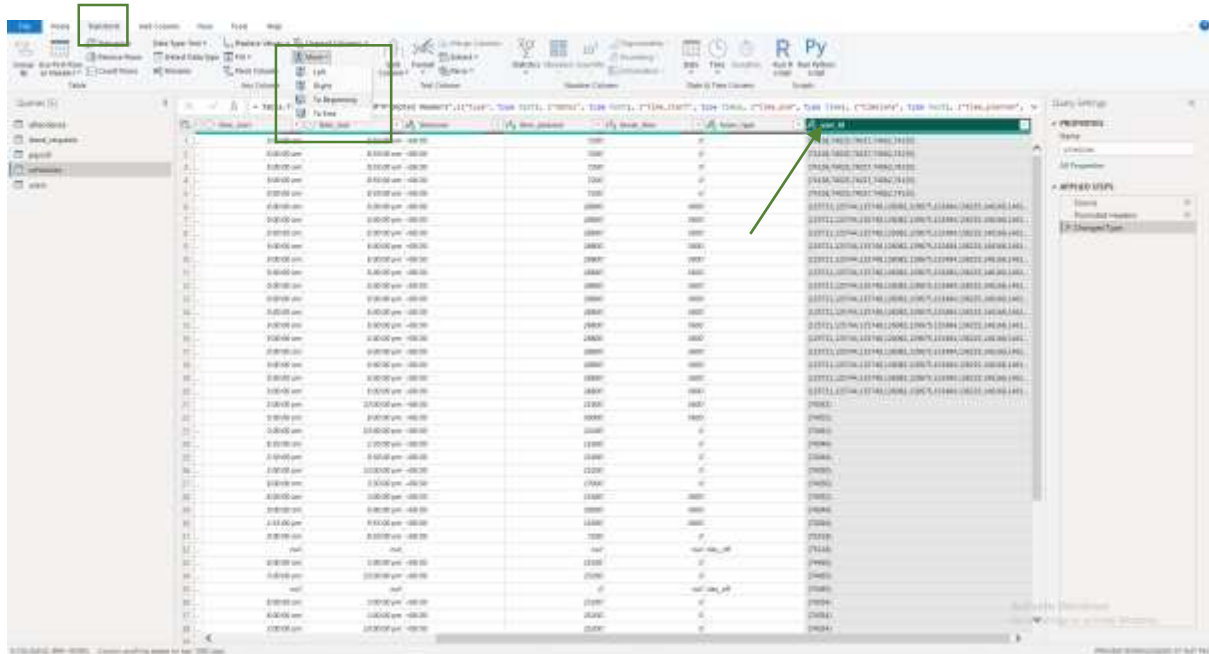


Result of populating the blank rows of currency column.

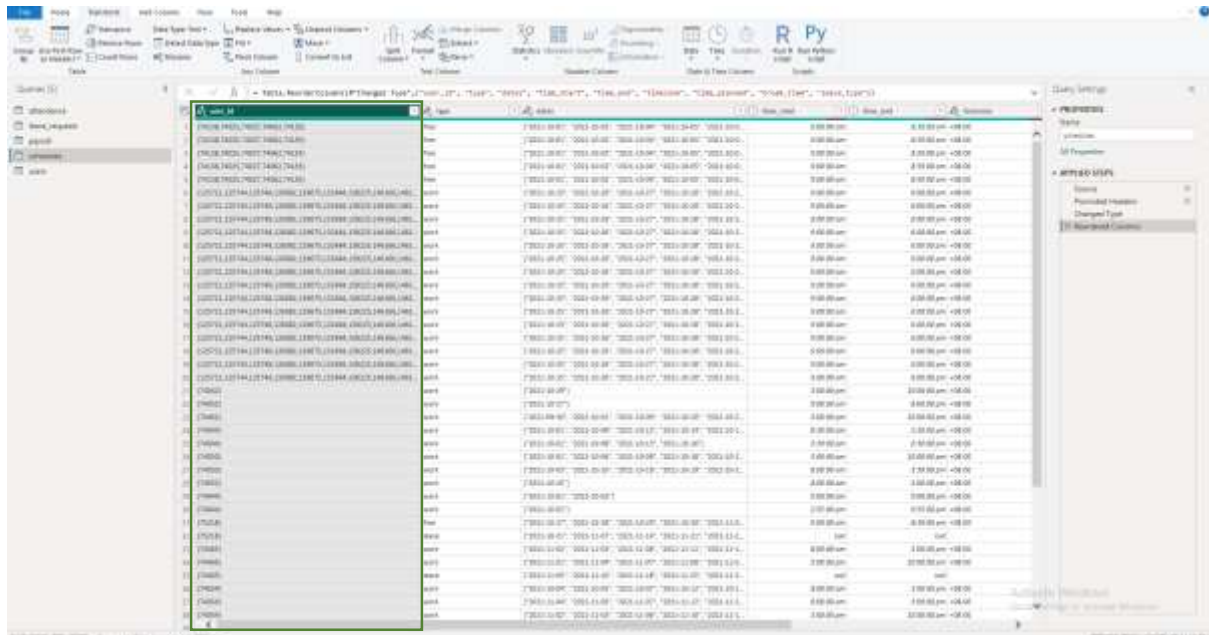


CLEANING AND TRANSFORMING SCHEDULES TABLE

Transferring the user_id column to the beginning of the table. Click to highlight the user_id column. Go to Transform, click the drop down arrow of Move then select To Beginning.

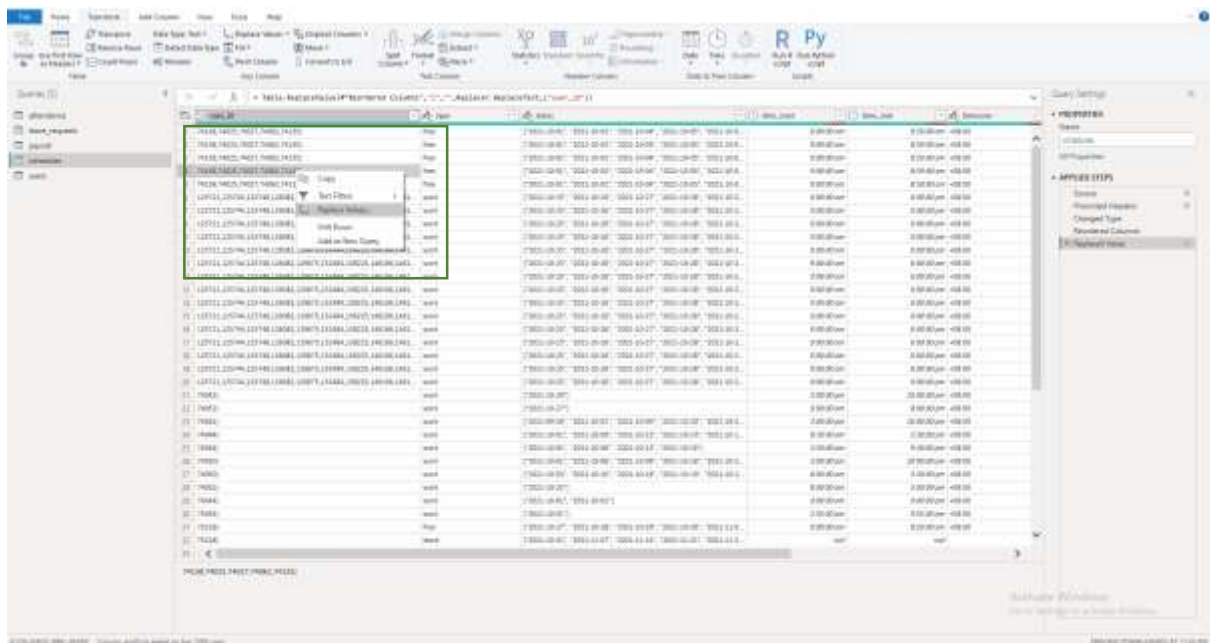


Result of schedules table column transfers of user_id.

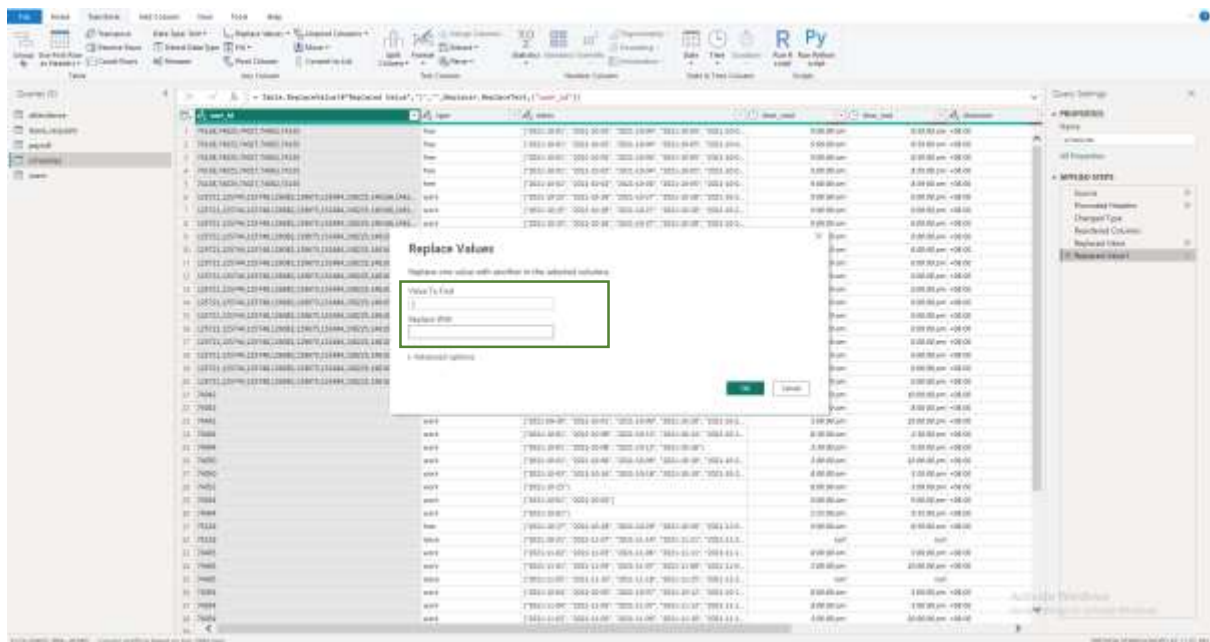


[illegible][illegible]

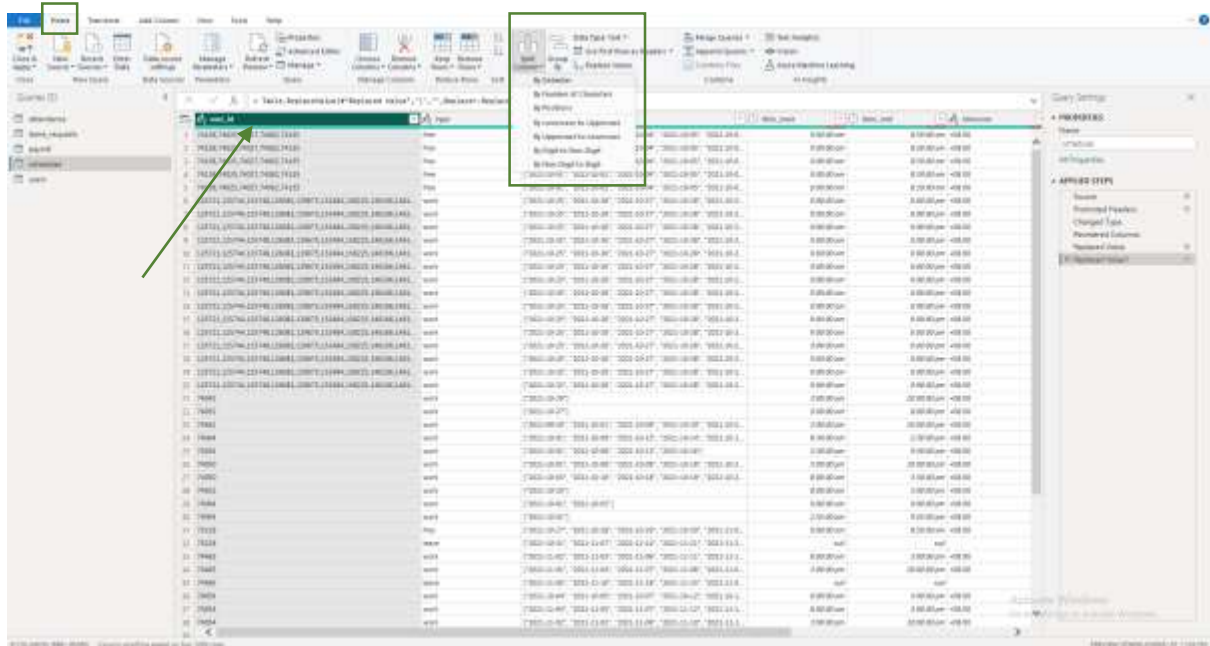
Right click again the user_id column cell and click the replace values.



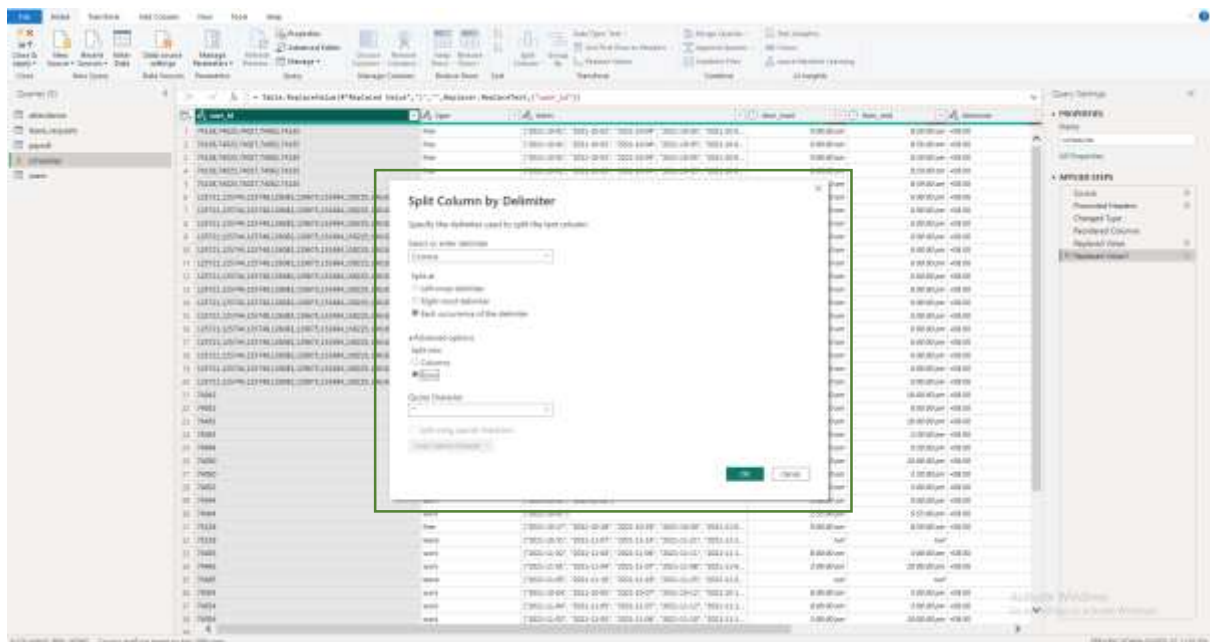
Input a close curly bracket to Values to Find, leave the Replace With input box blank and then click ok.



Splitting the value by delimiter of the user_id. Click to highlight the user_id column, go to Home and click the Split Column drop down arrow then select by Delimiter.

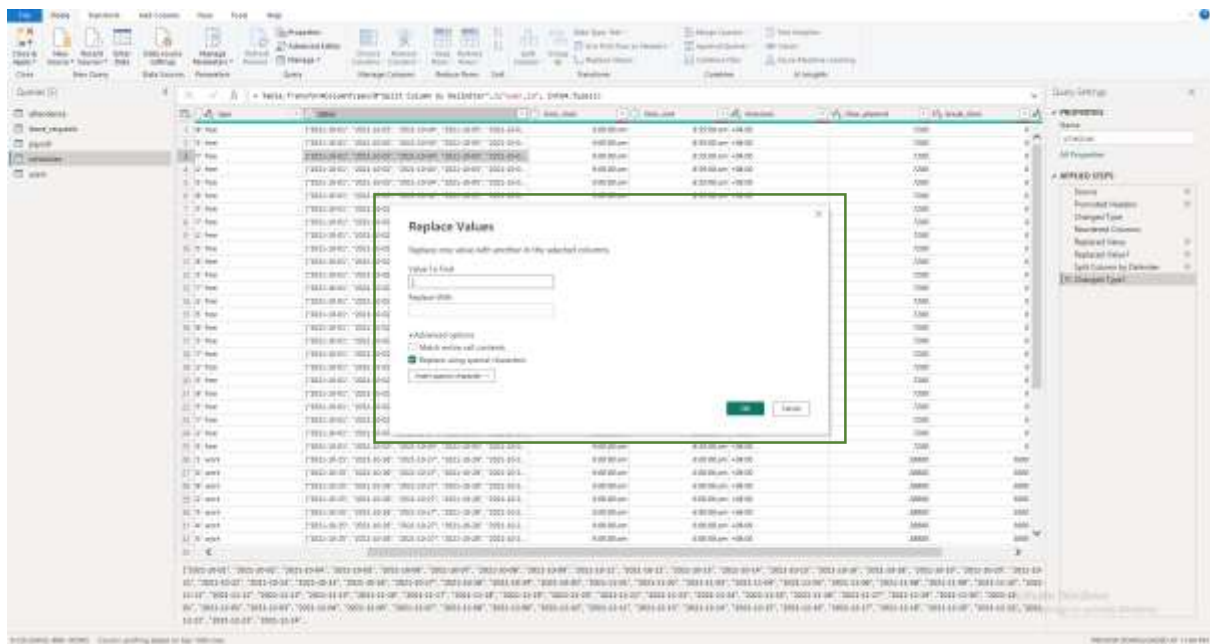


In the Split Column by Delimiter window, click the drop down arrow of Select or Enter delimiter and choose Comma. In the Split at choose Each occurrence of the delimiter, click Advance options and choose Rows and then click ok.

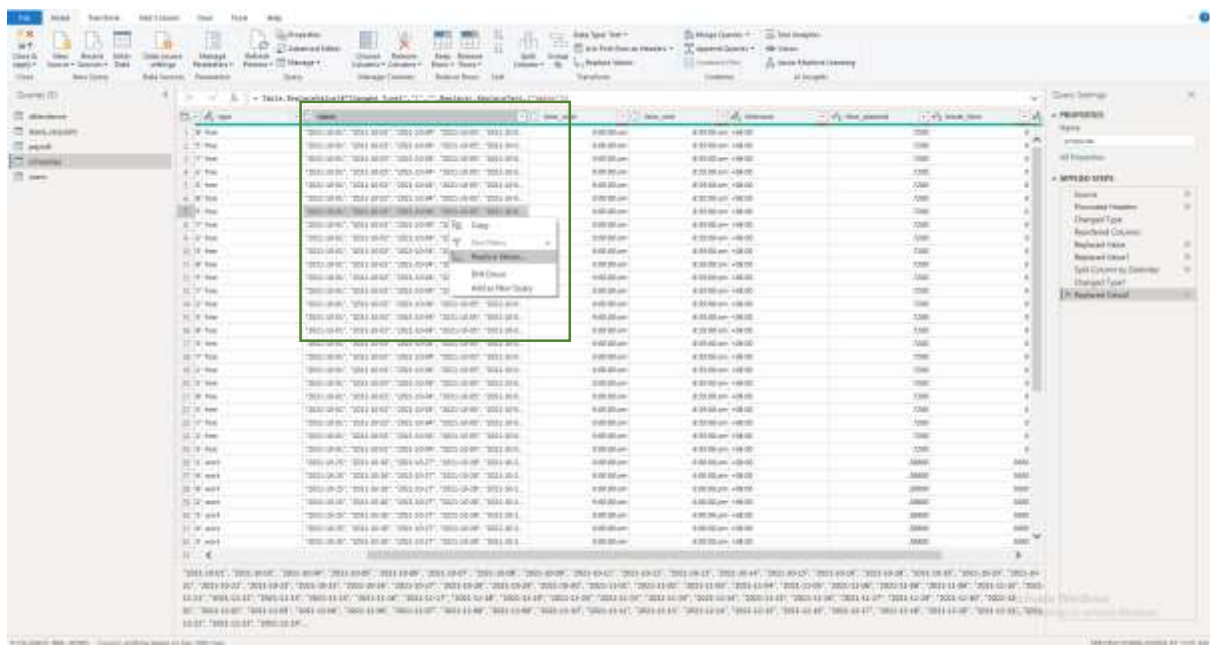


[illegible]

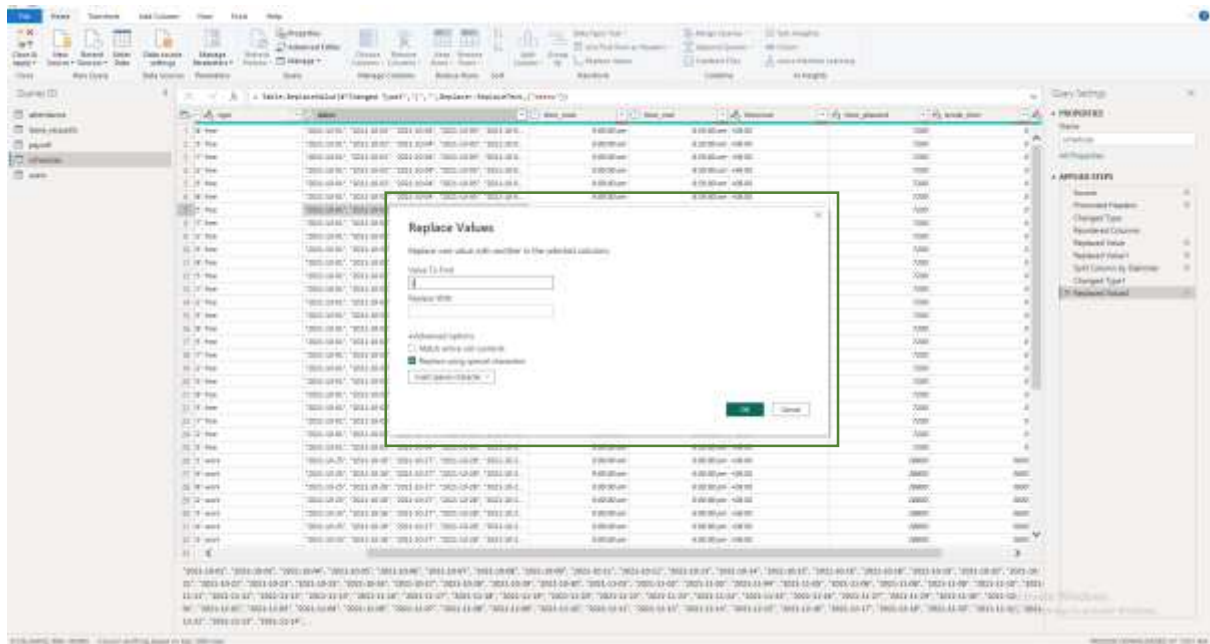
Input an open bracket to Values to Find, leave the Replace With input box blank and then click ok.



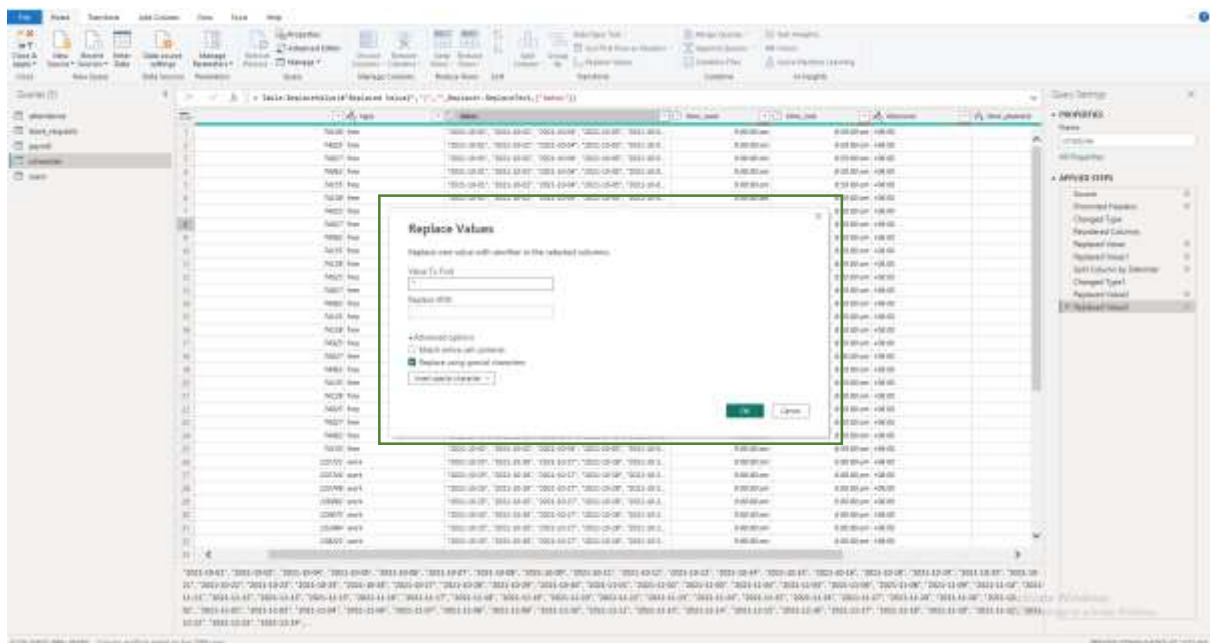
Right click again the cell of dates column and select replace values.



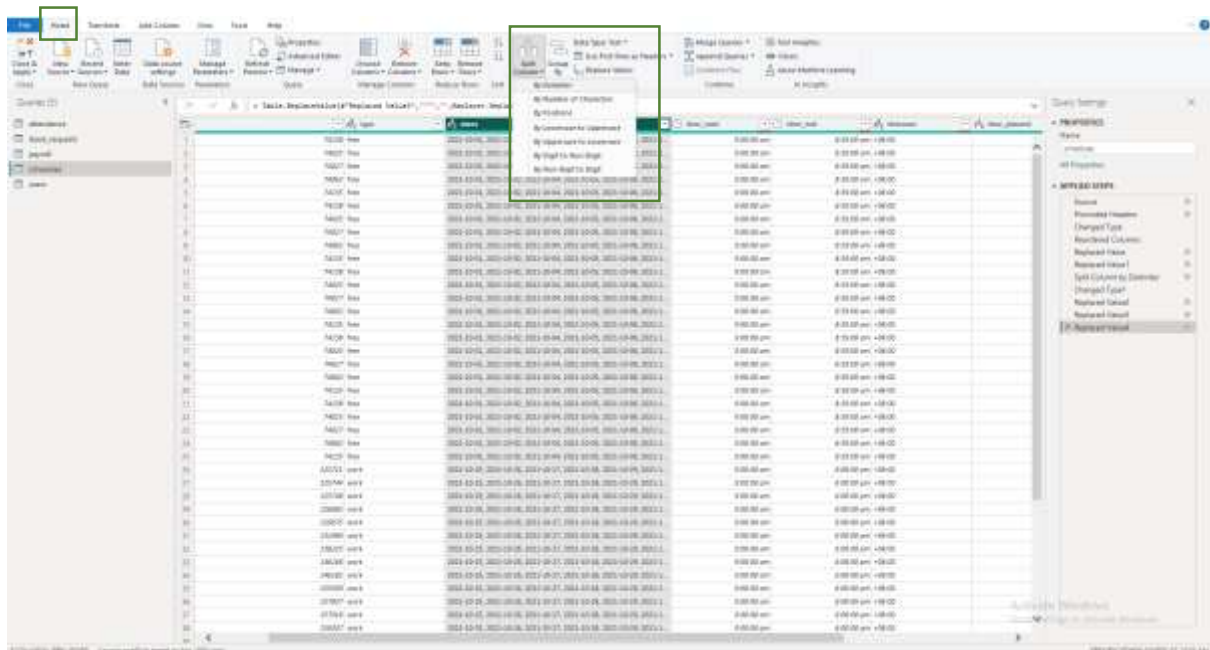
Input a close bracket to Values to Find, leave the Replace With input box blank and then click ok.



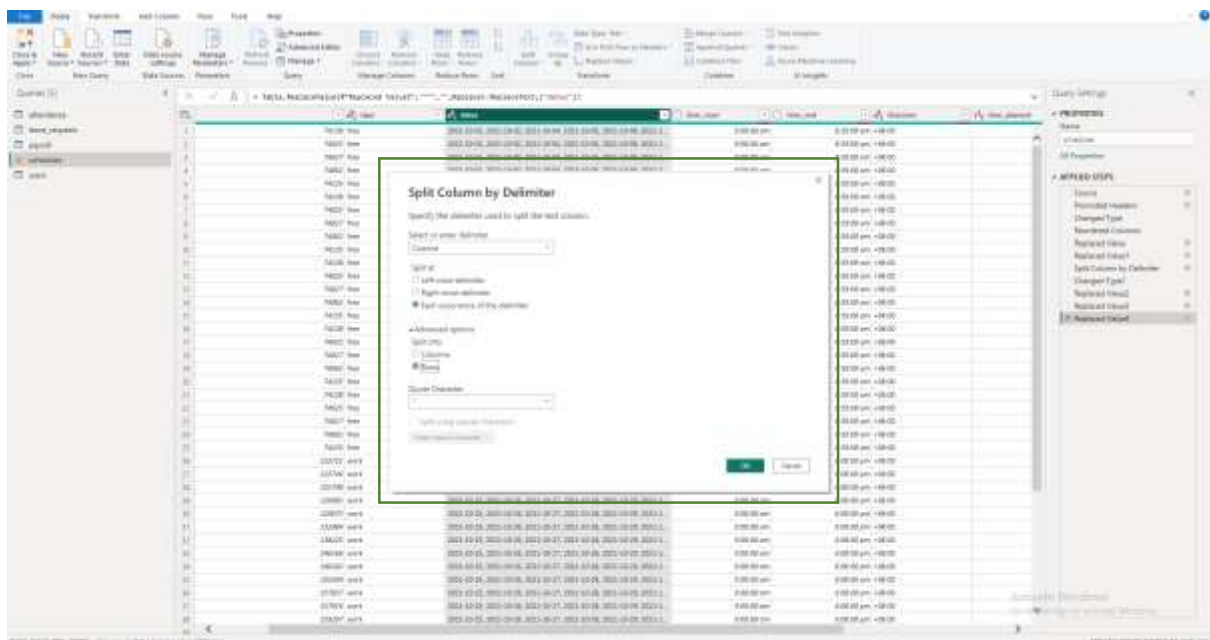
Right click again the cell of dates column and select replace values. Input an open double quote to Values to Find, leave the Replace With input box blank and then click ok.



Splitting the dates column by delimiter. Click to highlight the dates column, go to Home and click the Split Column dropdown arrow and select by Delimiter.



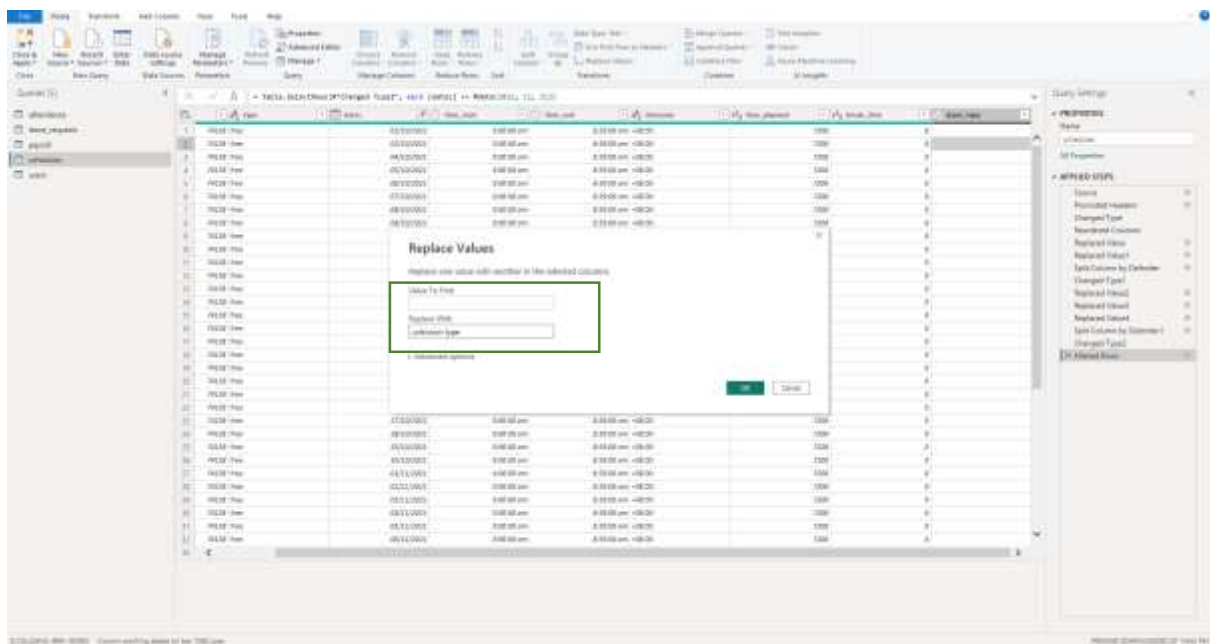
In the Split Column by Delimiter window, click the drop down arrow of Select or Enter delimiter and choose Comma. In the Split at choose Each occurrence of the delimiter, click Advance options and choose Rows and then click ok.



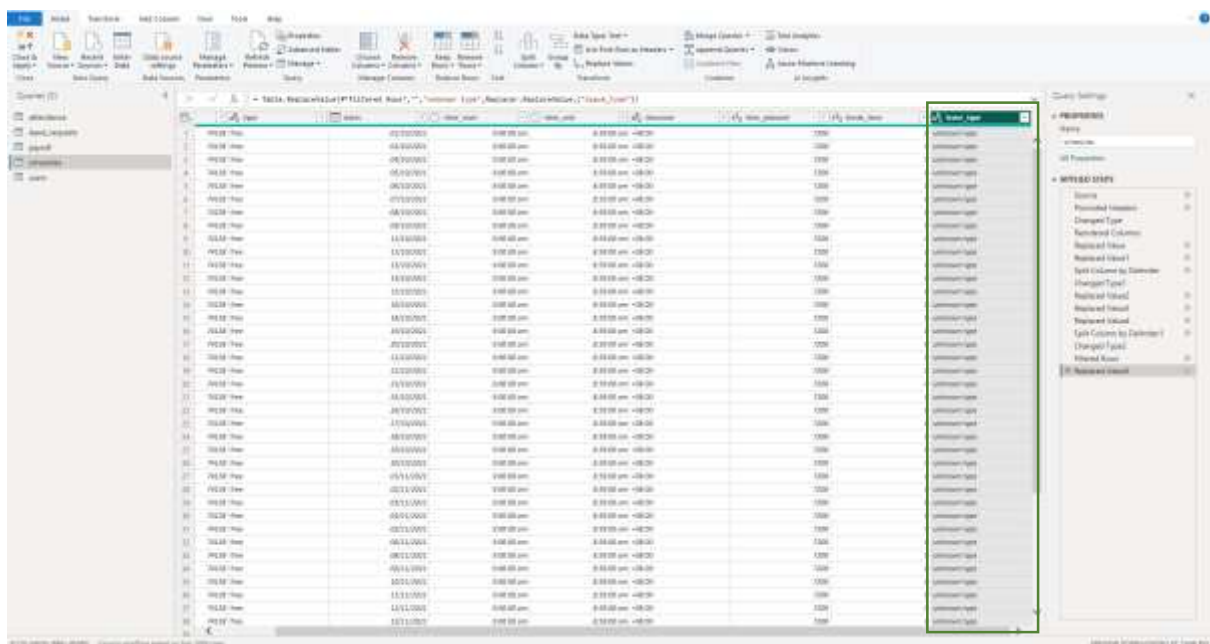
The screenshot displays the SAP S/4HANA Fiori 'Table Viewer' interface. The main table lists material stock data with columns: Material, Plant, Stock Type, Stock, and Stock Unit. A green box highlights the 'Stock' column. On the right, the 'Filter Settings' panel is open, showing a list of filters. A green arrow points to the 'Changed Stock' filter, which is currently selected.

The screenshot displays the Microsoft Access application window. At the top, the ribbon includes tabs for File, Home, Database Tools, and External Data. Below the ribbon, the title bar indicates the current database is 'Northwind.accdb'. The main workspace shows a query named 'qrySales' in Datasheet View. The query results are displayed in a grid with the following columns: Salesperson, Date Sold, Product, Price, Quantity, Total Price, and Tax Amount. The data rows show sales transactions for various products like 'Bottled Water', 'Energy Drink', 'Soft Drink', etc., across different dates from 06/07/2010 to 08/11/2010. A right-click context menu is open over the 'Date Sold' column header, offering actions such as Copy, Paste, Cut, Replace Values, Sort Ascending, Sort Descending, Add to Filter/Query, and Show/Hide Fields.

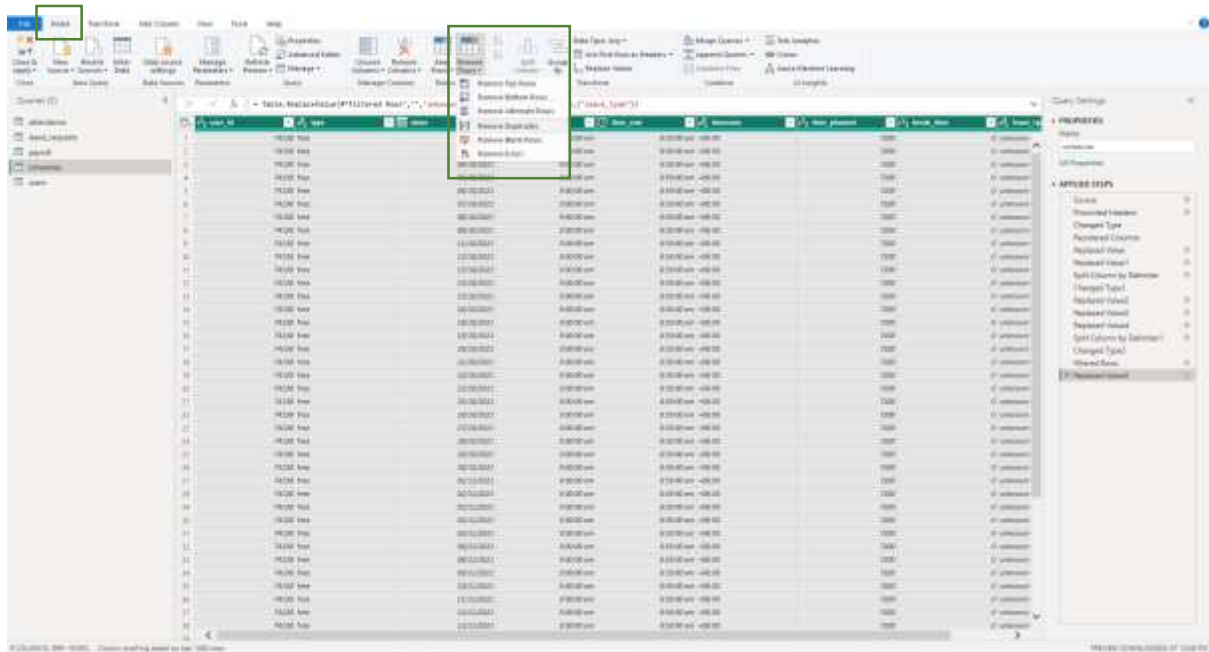
Leave the input box of Value to Find blank, input unknown type to the Replace With input box and then click ok.



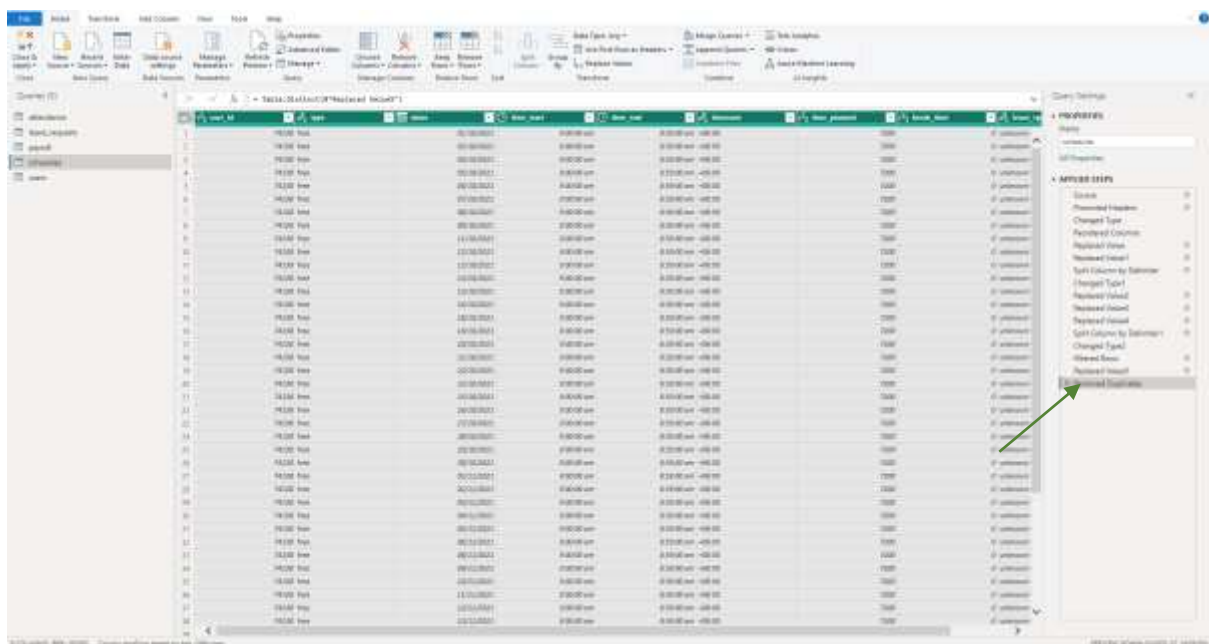
Result of replacing blank value in the leaf_type column with unknown type



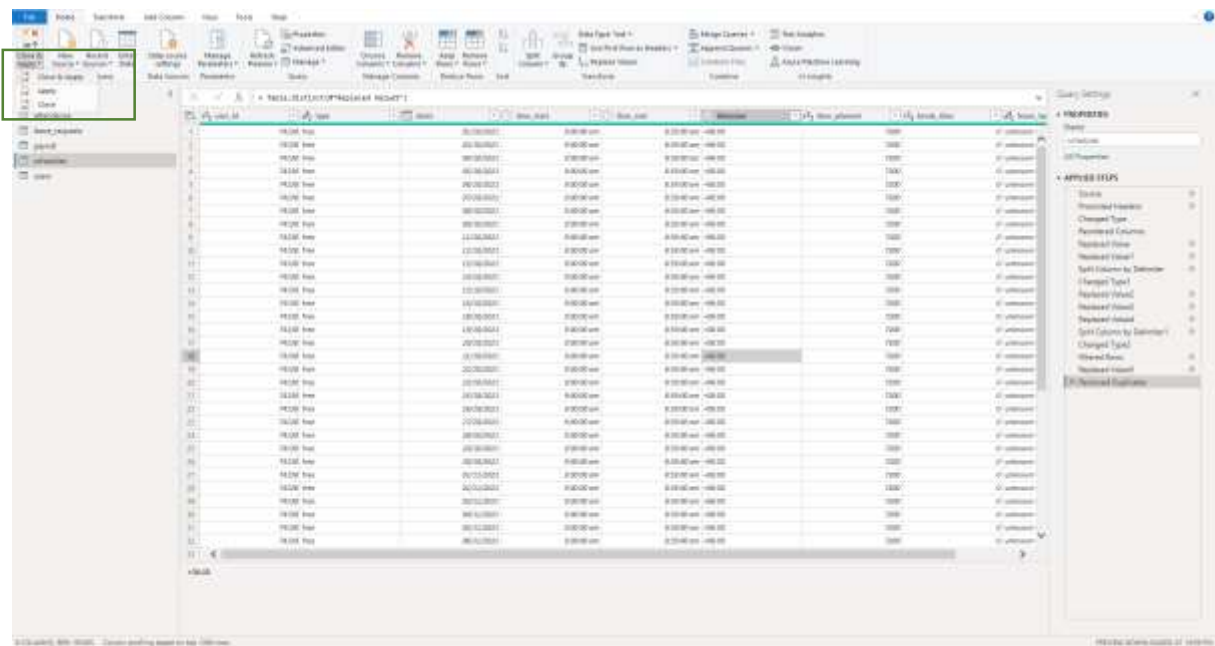
Removing duplicate values in schedules table. Highlight all the column of schedules table. Go to Home, select and click the drop down arrow of Remove Rows and then click Remove Duplicates.



Result of removing duplicate values of schedules table

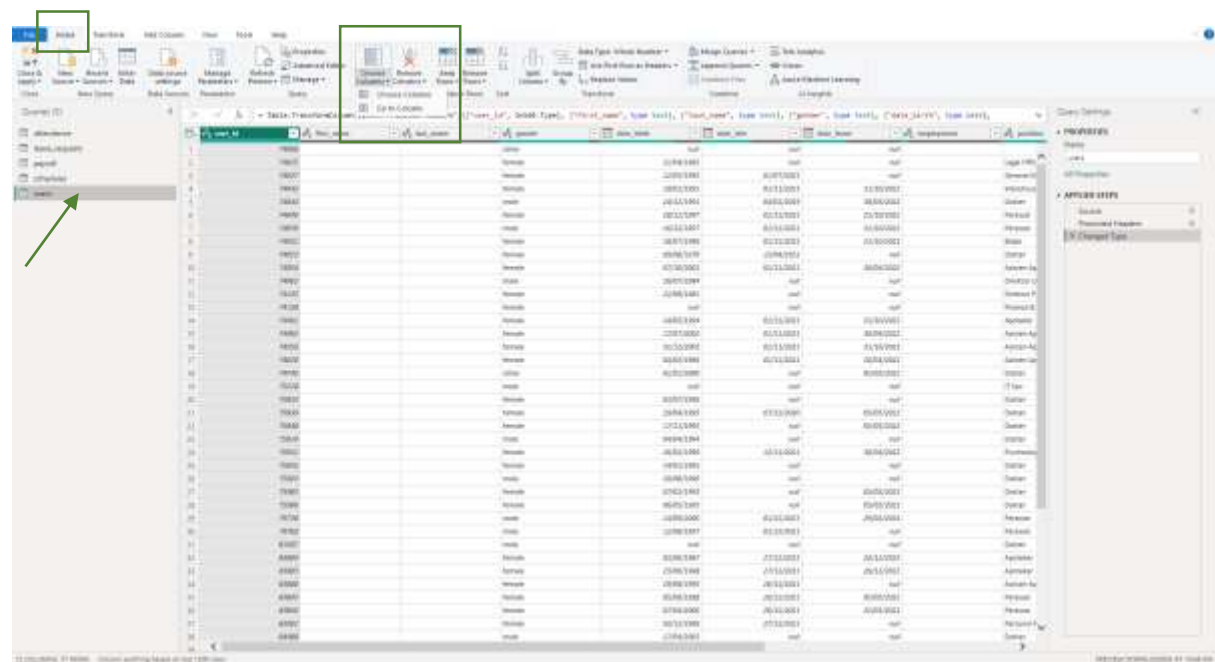


Total of 21676 rows remain in the schedule table after removing duplicate rows. Click close and apply to load the data.

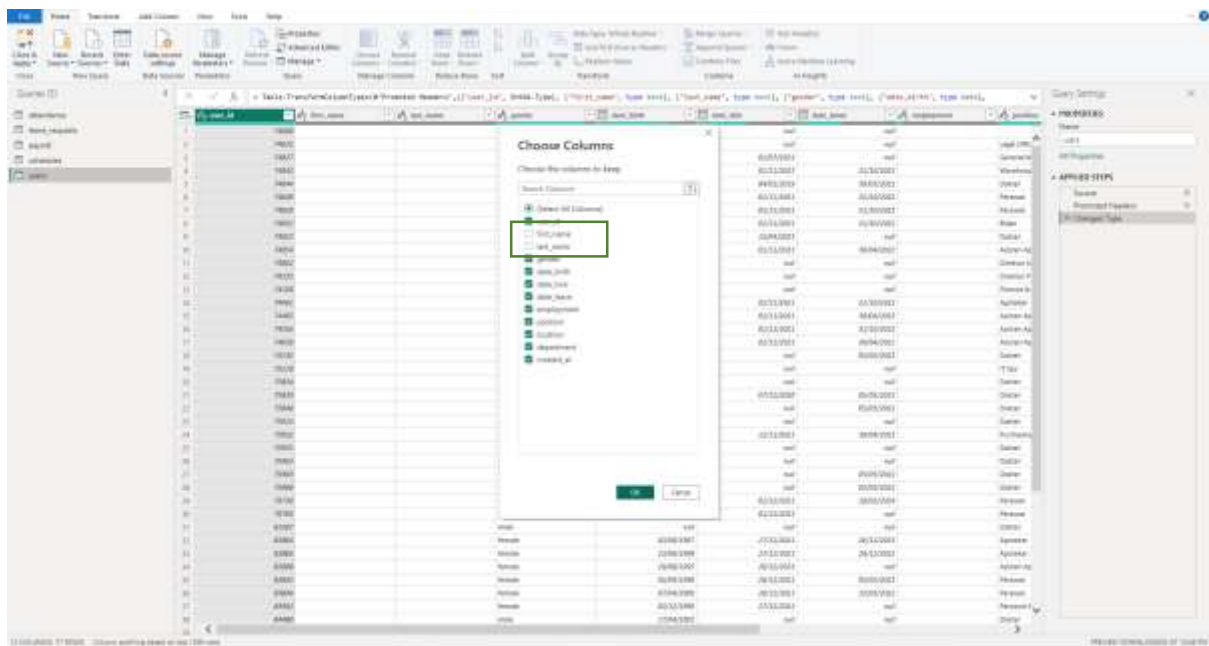


CLEANING AND TRANSFORMING USERS TABLE

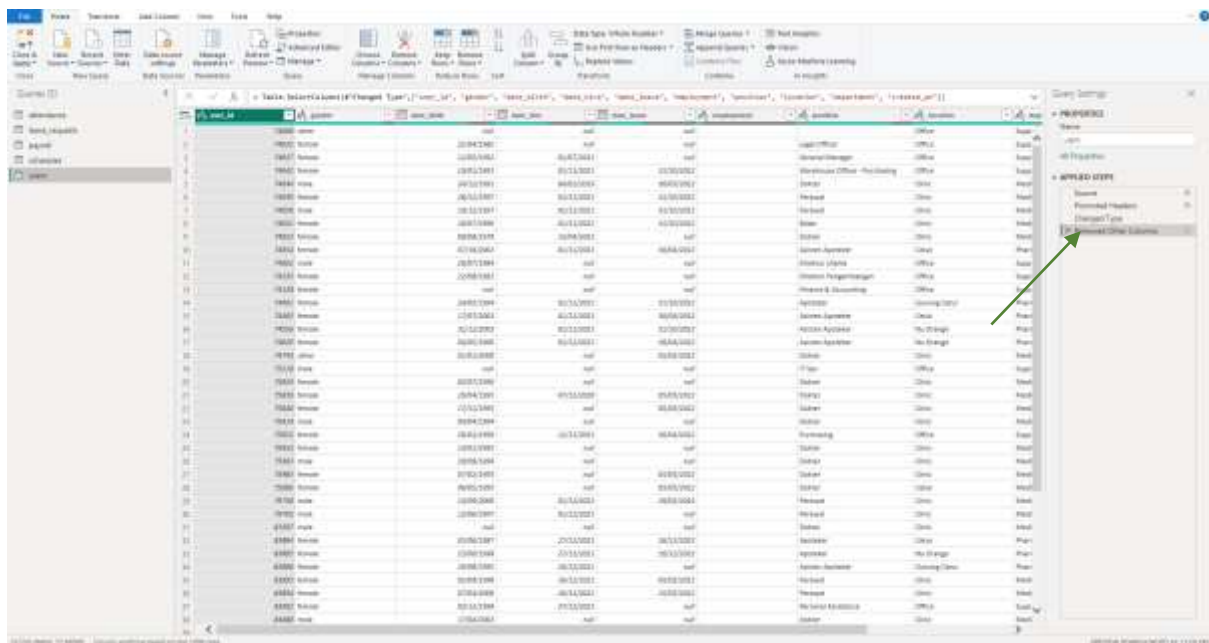
Removing unwanted column in users table. Select the users table in the Navigator Pane. Go to Home, click the drop down arrow of Choose Columns and click Choose Columns.



Uncheck the box of column you want to remove from the table, we choose first_name and last_name and then click ok.



Result of removing unwanted columns from the users table.



The screenshot shows the 'Replace Values' dialog box in Microsoft Excel. The dialog is open over a spreadsheet. The 'Find what:' field contains 'Salesperson' and the 'Replace with:' field contains 'Product'. The 'Advanced options' section is expanded, showing 'Match entire cell contents' and 'Match cell content with formulas' checked. The 'Replace' button is highlighted.

Result of translating the position column values into English.

No	NIK	NAME	position	APPROVED
1	12.000.0000	PERAWAT	Nurse	
2	12.000.0001	PERAWAT	Nurse	
3	12.000.0002	PERAWAT	Nurse	
4	12.000.0003	PERAWAT	Nurse	
5	12.000.0004	PERAWAT	Nurse	
6	12.000.0005	PERAWAT	Nurse	
7	12.000.0006	PERAWAT	Nurse	
8	12.000.0007	PERAWAT	Nurse	
9	12.000.0008	PERAWAT	Nurse	
10	12.000.0009	PERAWAT	Nurse	
11	12.000.0010	PERAWAT	Nurse	
12	12.000.0011	PERAWAT	Nurse	
13	12.000.0012	PERAWAT	Nurse	
14	12.000.0013	PERAWAT	Nurse	
15	12.000.0014	PERAWAT	Nurse	
16	12.000.0015	PERAWAT	Nurse	
17	12.000.0016	PERAWAT	Nurse	
18	12.000.0017	PERAWAT	Nurse	
19	12.000.0018	PERAWAT	Nurse	
20	12.000.0019	PERAWAT	Nurse	
21	12.000.0020	PERAWAT	Nurse	
22	12.000.0021	PERAWAT	Nurse	
23	12.000.0022	PERAWAT	Nurse	
24	12.000.0023	PERAWAT	Nurse	
25	12.000.0024	PERAWAT	Nurse	
26	12.000.0025	PERAWAT	Nurse	
27	12.000.0026	PERAWAT	Nurse	
28	12.000.0027	PERAWAT	Nurse	
29	12.000.0028	PERAWAT	Nurse	
30	12.000.0029	PERAWAT	Nurse	
31	12.000.0030	PERAWAT	Nurse	
32	12.000.0031	PERAWAT	Nurse	
33	12.000.0032	PERAWAT	Nurse	
34	12.000.0033	PERAWAT	Nurse	
35	12.000.0034	PERAWAT	Nurse	
36	12.000.0035	PERAWAT	Nurse	
37	12.000.0036	PERAWAT	Nurse	
38	12.000.0037	PERAWAT	Nurse	
39	12.000.0038	PERAWAT	Nurse	
40	12.000.0039	PERAWAT	Nurse	
41	12.000.0040	PERAWAT	Nurse	
42	12.000.0041	PERAWAT	Nurse	
43	12.000.0042	PERAWAT	Nurse	
44	12.000.0043	PERAWAT	Nurse	
45	12.000.0044	PERAWAT	Nurse	
46	12.000.0045	PERAWAT	Nurse	
47	12.000.0046	PERAWAT	Nurse	
48	12.000.0047	PERAWAT	Nurse	
49	12.000.0048	PERAWAT	Nurse	
50	12.000.0049	PERAWAT	Nurse	
51	12.000.0050	PERAWAT	Nurse	
52	12.000.0051	PERAWAT	Nurse	
53	12.000.0052	PERAWAT	Nurse	
54	12.000.0053	PERAWAT	Nurse	
55	12.000.0054	PERAWAT	Nurse	
56	12.000.0055	PERAWAT	Nurse	
57	12.000.0056	PERAWAT	Nurse	
58	12.000.0057	PERAWAT	Nurse	
59	12.000.0058	PERAWAT	Nurse	
60	12.000.0059	PERAWAT	Nurse	
61	12.000.0060	PERAWAT	Nurse	
62	12.000.0061	PERAWAT	Nurse	
63	12.000.0062	PERAWAT	Nurse	
64	12.000.0063	PERAWAT	Nurse	
65	12.000.0064	PERAWAT	Nurse	
66	12.000.0065	PERAWAT	Nurse	
67	12.000.0066	PERAWAT	Nurse	
68	12.000.0067	PERAWAT	Nurse	
69	12.000.0068	PERAWAT	Nurse	
70	12.000.0069	PERAWAT	Nurse	
71	12.000.0070	PERAWAT	Nurse	
72	12.000.0071	PERAWAT	Nurse	
73	12.000.0072	PERAWAT	Nurse	
74	12.000.0073	PERAWAT	Nurse	
75	12.000.0074	PERAWAT	Nurse	
76	12.000.0075	PERAWAT	Nurse	
77	12.000.0076	PERAWAT	Nurse	
78	12.000.0077	PERAWAT	Nurse	
79	12.000.0078	PERAWAT	Nurse	
80	12.000.0079	PERAWAT	Nurse	
81	12.000.0080	PERAWAT	Nurse	
82	12.000.0081	PERAWAT	Nurse	
83	12.000.0082	PERAWAT	Nurse	
84	12.000.0083	PERAWAT	Nurse	
85	12.000.0084	PERAWAT	Nurse	
86	12.000.0085	PERAWAT	Nurse	
87	12.000.0086	PERAWAT	Nurse	
88	12.000.0087	PERAWAT	Nurse	
89	12.000.0088	PERAWAT	Nurse	
90	12.000.0089	PERAWAT	Nurse	
91	12.000.0090	PERAWAT	Nurse	
92	12.000.0091	PERAWAT	Nurse	
93	12.000.0092	PERAWAT	Nurse	
94	12.000.0093	PERAWAT	Nurse	
95	12.000.0094	PERAWAT	Nurse	
96	12.000.0095	PERAWAT	Nurse	
97	12.000.0096	PERAWAT	Nurse	
98	12.000.0097	PERAWAT	Nurse	
99	12.000.0098	PERAWAT	Nurse	
100	12.000.0099	PERAWAT	Nurse	

Repeat the same procedure to replace all the position values into English.

Values to Find

Perawat

Bidan

Asisten Apoteker

Direktur Utama

Direktur Pengembangan

Apoteker

Admin Bisnis

IT Spv

Blank

Replace With

nurse

midwife

pharmacist assistant

president director

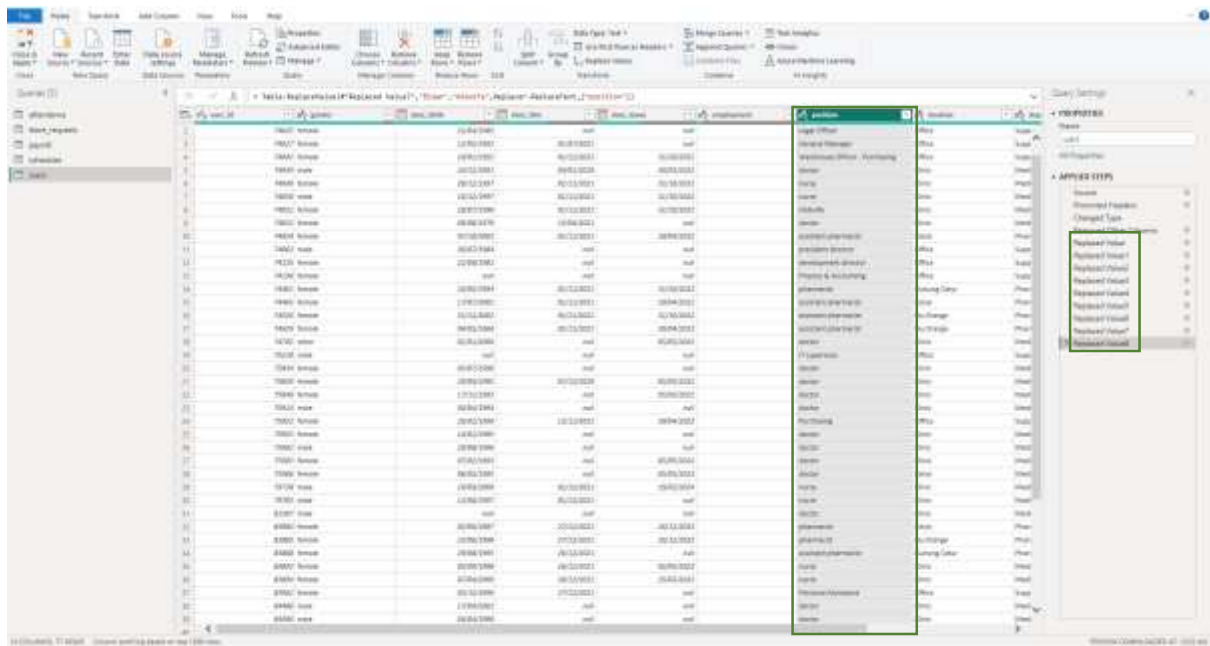
development director

pharmacist

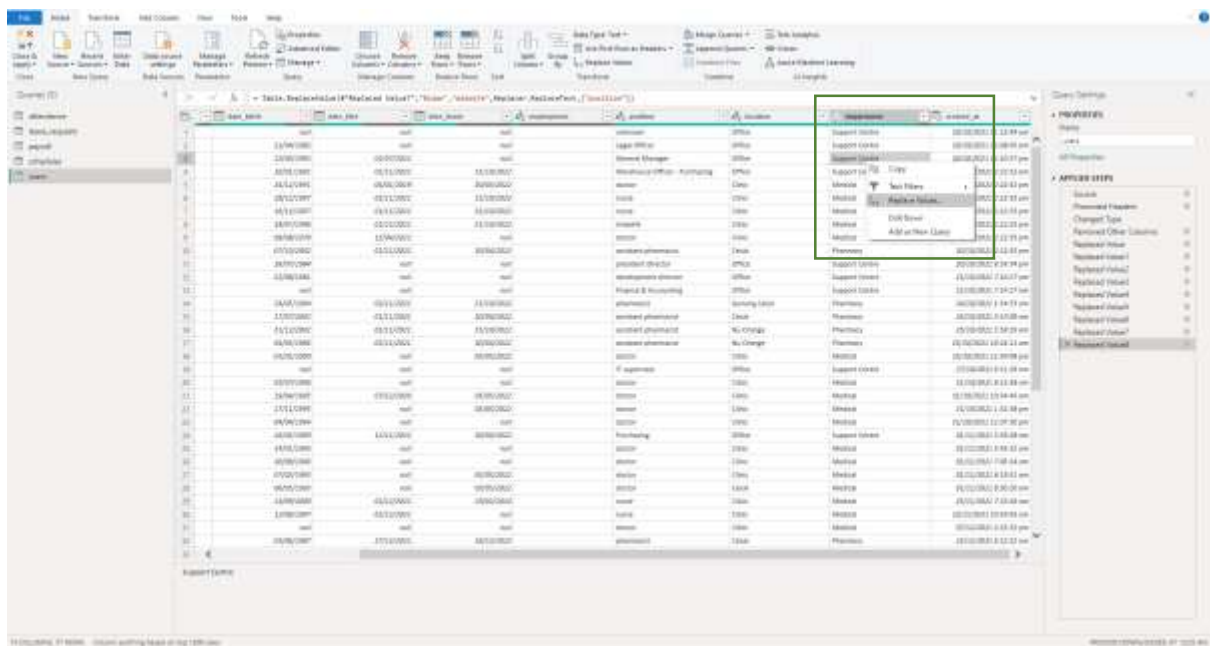
business admins

IT supervisor

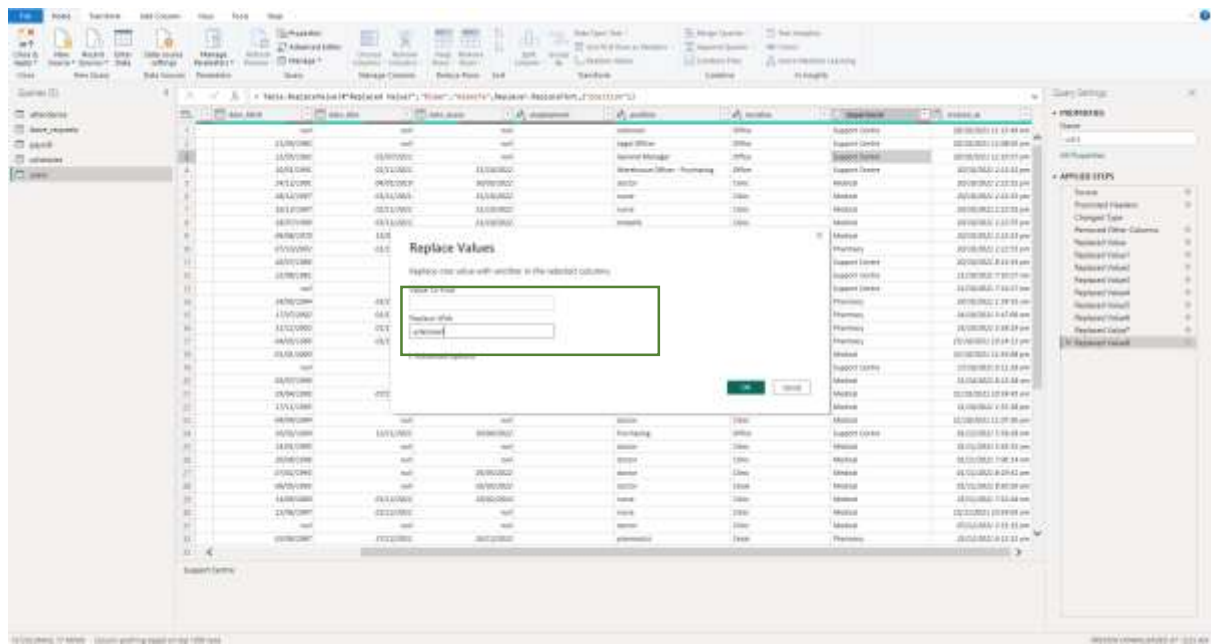
unknown



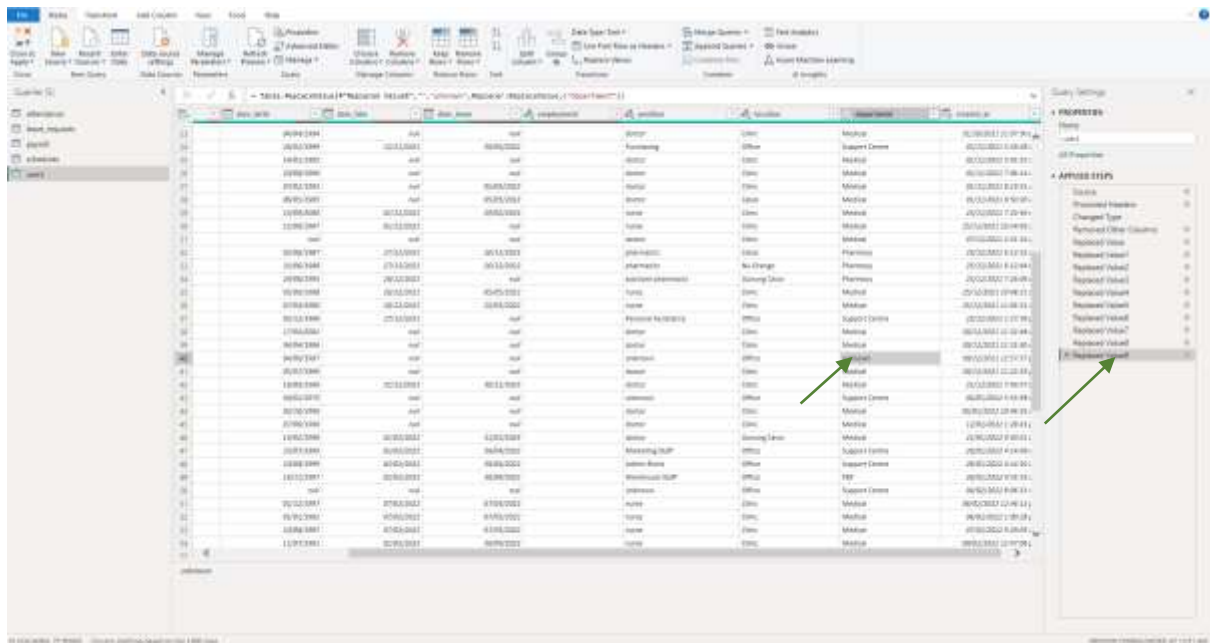
Replacing the blank value of department column. Right click the cell of the department column, click replace values.



Leave the Value to Find blank, input unknown to Replace With input box and then click ok.



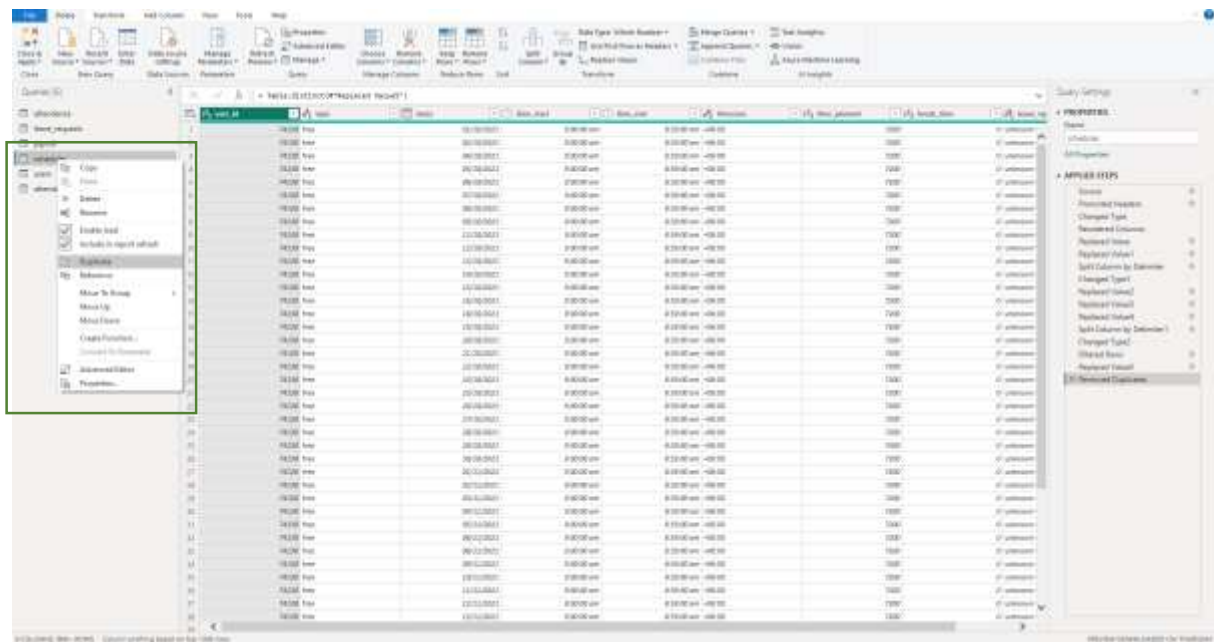
Result of replacing blank value in department column. Click Close to apply to load the data to model.



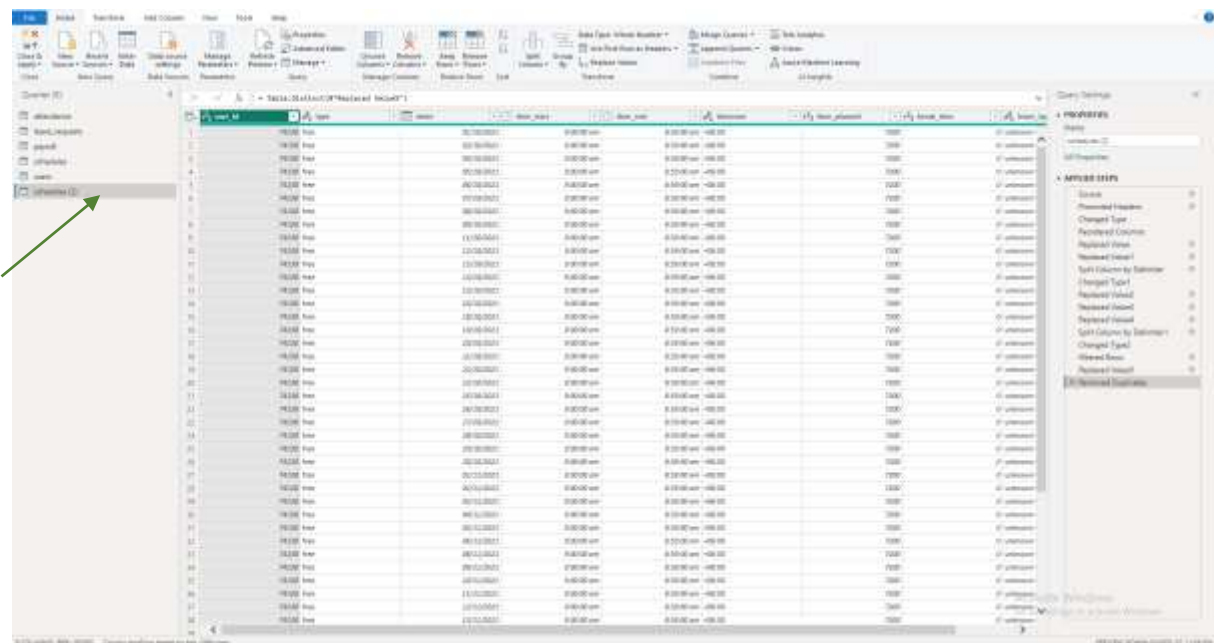
The cleaning and transformation of the five tables is finished. The next step is to merge the schedules table with attendance table.

MERGING ATTENDANCE AND SCHEDULES TABLE

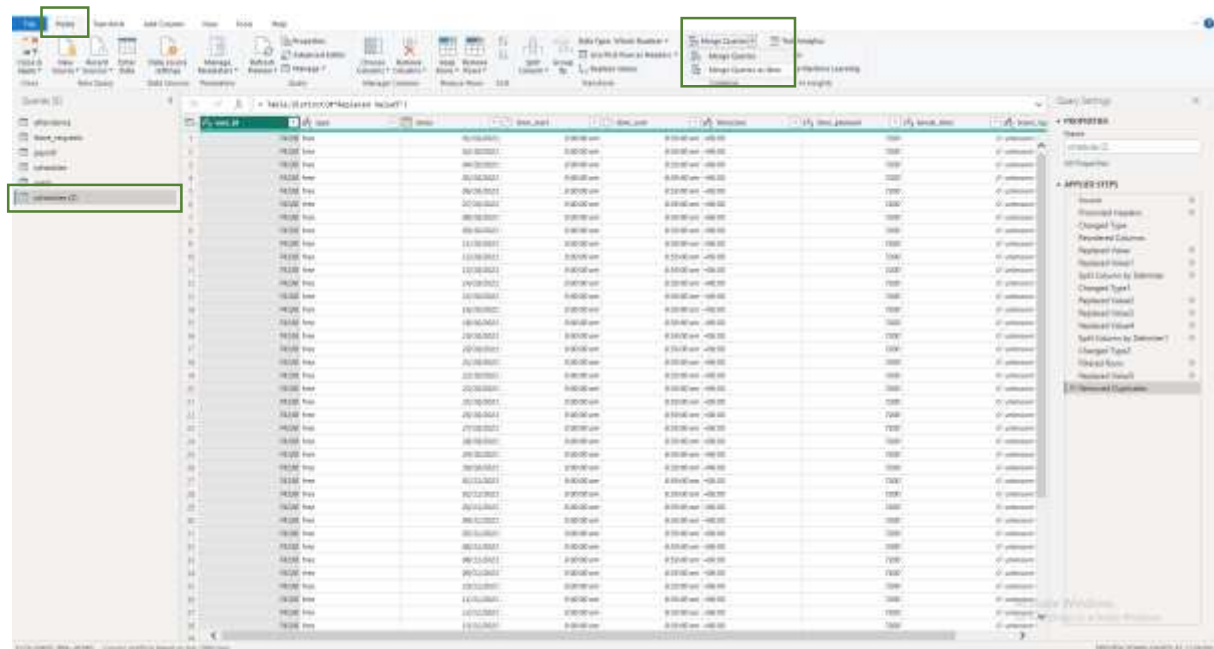
Make a duplicate of the schedules table by Right Clicking the schedules table to the Navigator Pane and click Duplicate.



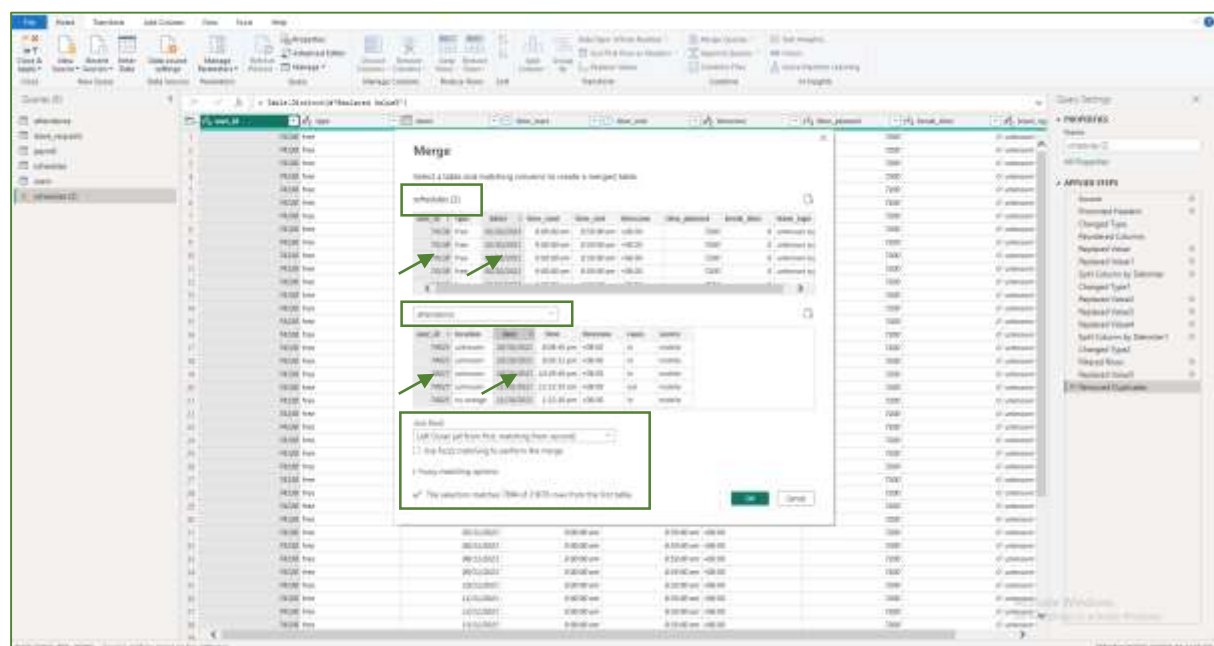
Results of duplicating the schedules table.



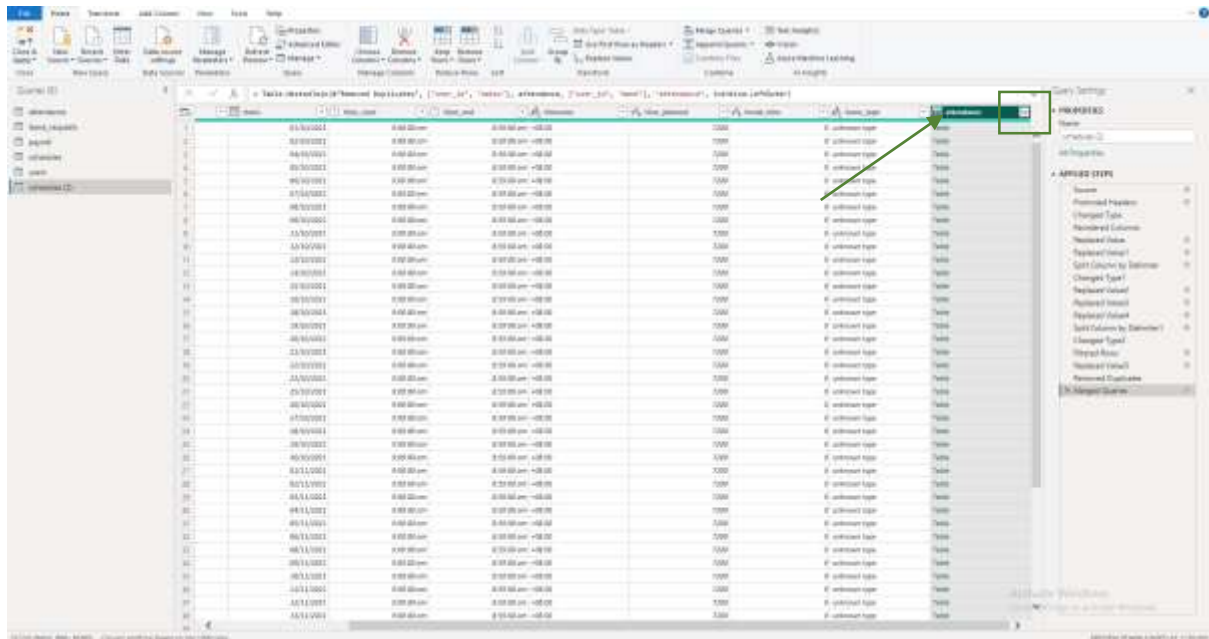
Click the duplicate of schedules table in the Navigator Pane, go to Home and click the drop down arrow of Merge Queries and then click the Merge Queries.



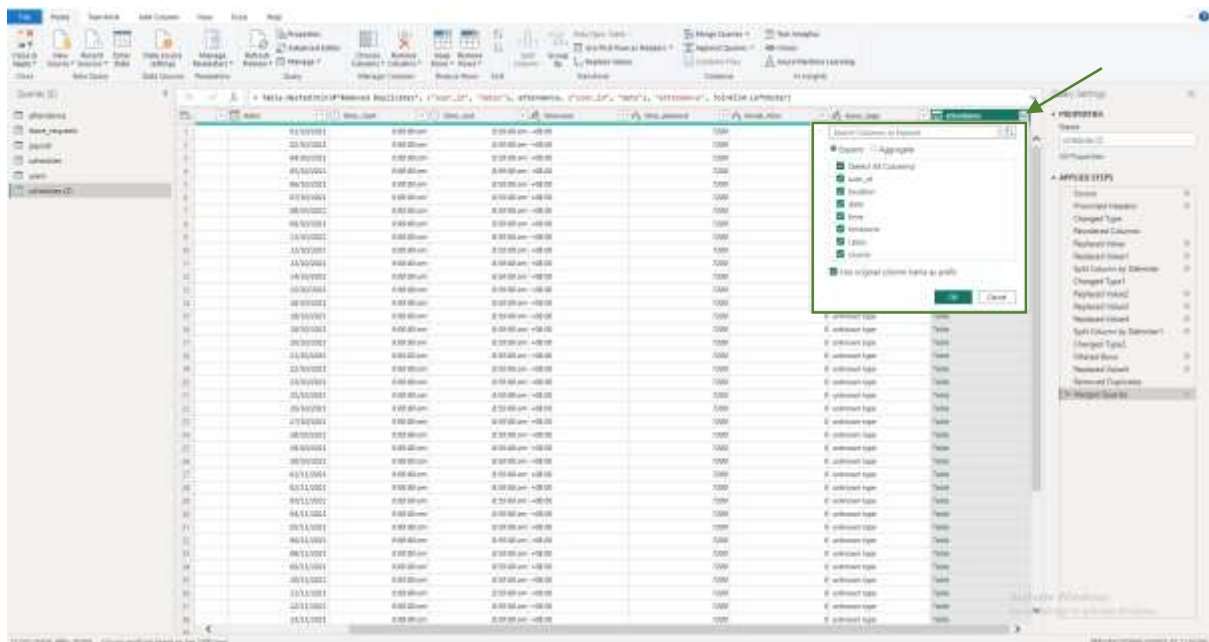
The Merge pop up window appear, this is the format we did when setting the Merge Queries. The first table is the duplicate of the schedules table then the second table is the attendance. The commonality of these two tables is the user_id and dates. Then the Join Kind used is Left Outer. Click ok to continue the process.



A new column will appear, for now it is still compressed. To display the entire table attendance. Click the icon on the side of the column attendance.



After clicking the icon next to attendance, these check boxes will appear. Here you can select the desired column to appear in the attendance table. Let's leave all columns checked for now.



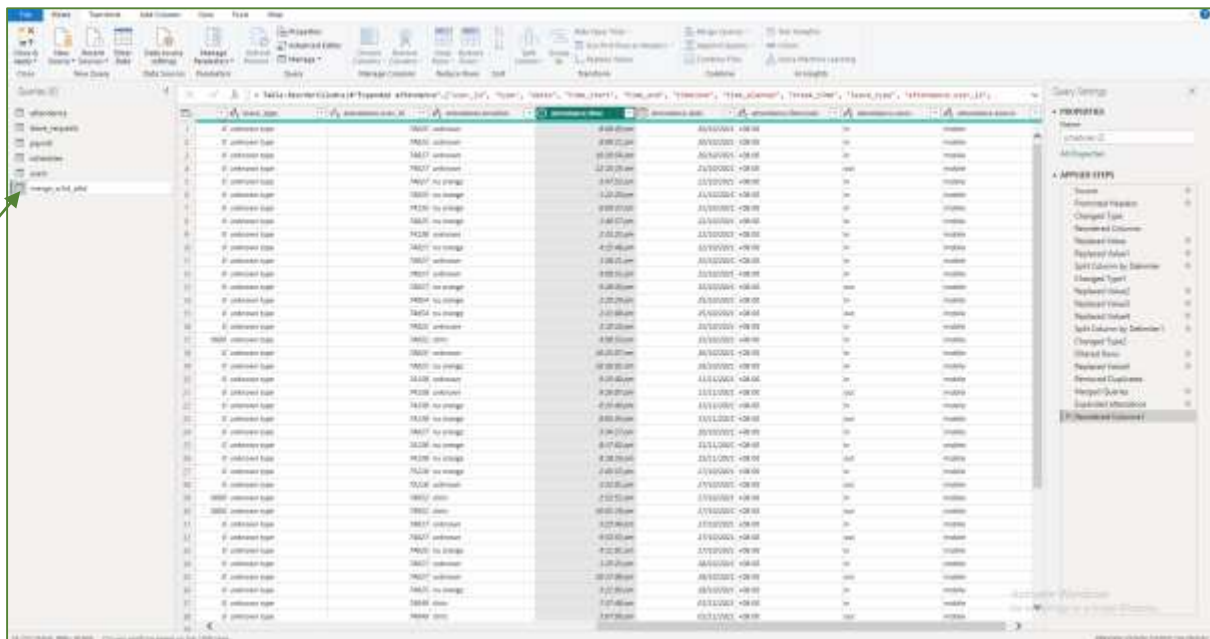
This is the result of merging schedules and attendance table.

The screenshot shows a software interface with a table containing merged schedule and attendance data. The table has columns for ID, Name, Date, Time, and Status. A green arrow points to the 'Status' column, which contains values like 'Present', 'Absent', and 'No Show'. The right sidebar shows a 'Query Settings' panel with a 'Table' dropdown set to 'Attendance' and a 'Columns' list including 'ID', 'Name', 'Date', 'Time', and 'Status'.

Renaming the new merge table schedules (2) table with merge_schd_attd. Right click schedules (2) in Navigator Pane, click rename.

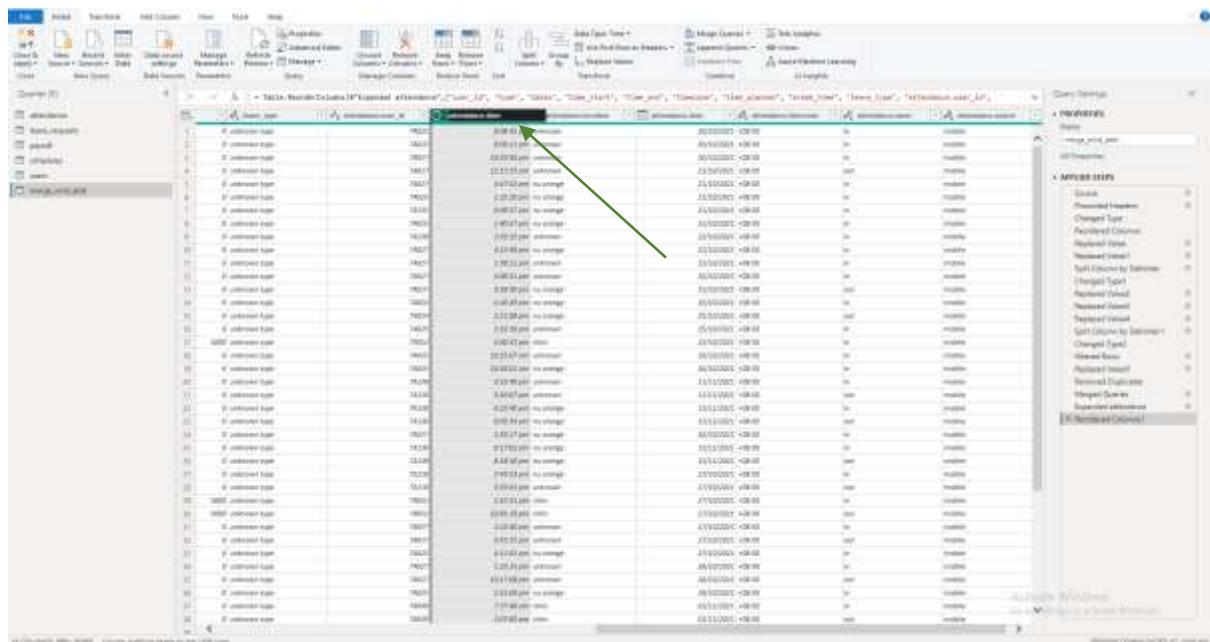
The screenshot shows a software interface with a table containing merged schedule and attendance data. A green box highlights the 'Table' dropdown in the 'Query Settings' panel, which is set to 'Attendance'. A green arrow points to the 'Table' dropdown, which is set to 'Attendance'. The table has columns for ID, Name, Date, Time, and Status. The right sidebar shows a 'Query Settings' panel with a 'Table' dropdown set to 'Attendance' and a 'Columns' list including 'ID', 'Name', 'Date', 'Time', and 'Status'.

Change the table name to merge_schd_attd.



CLEANING AND TRANSFORMING MERGE_SCHD_ATT

We want to move the position of the attendance.time column and place it next to time_end. Click and hold the attendance.time column and drag it next to time_end.



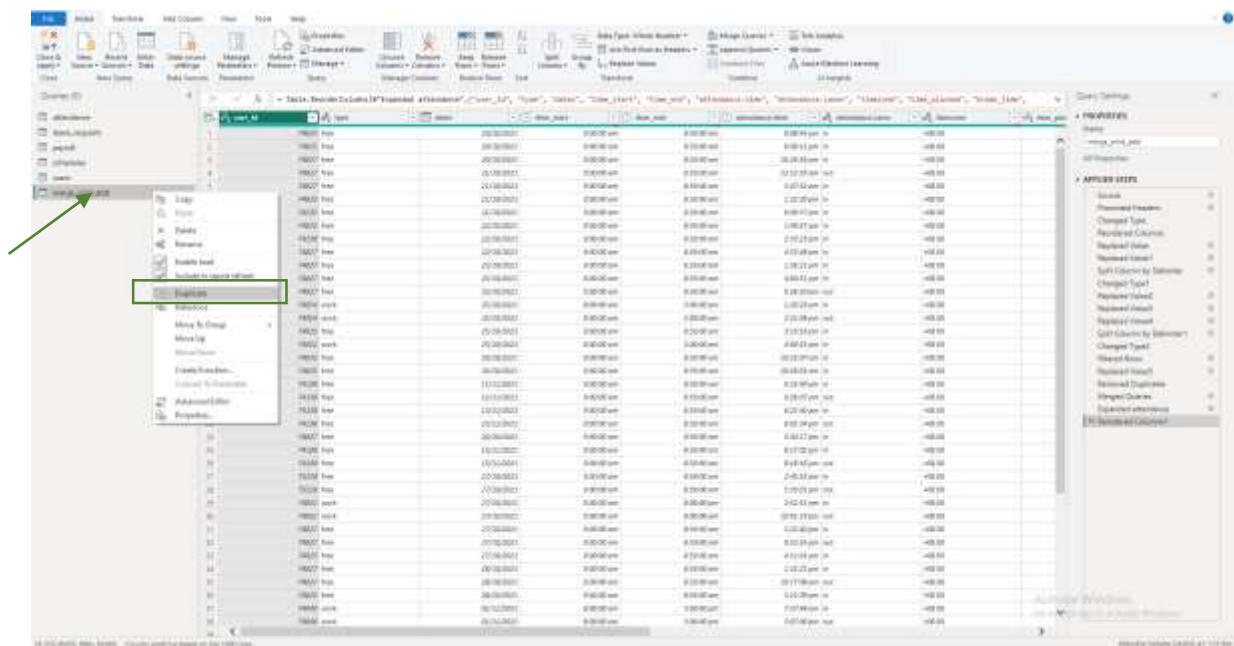
We also want to move the position of the attendance.cases column and put it next to the attendance.time column. Click and hold the attendance.cases column and then drag it next to the attendance.time.

The screenshot shows a data management interface with a table containing multiple columns. The columns are labeled: attendance.time, attendance.cases, attendance.time, attendance.cases, attendance.time, attendance.cases. A green arrow points to the 'attendance.cases' column header, indicating the action of moving it.

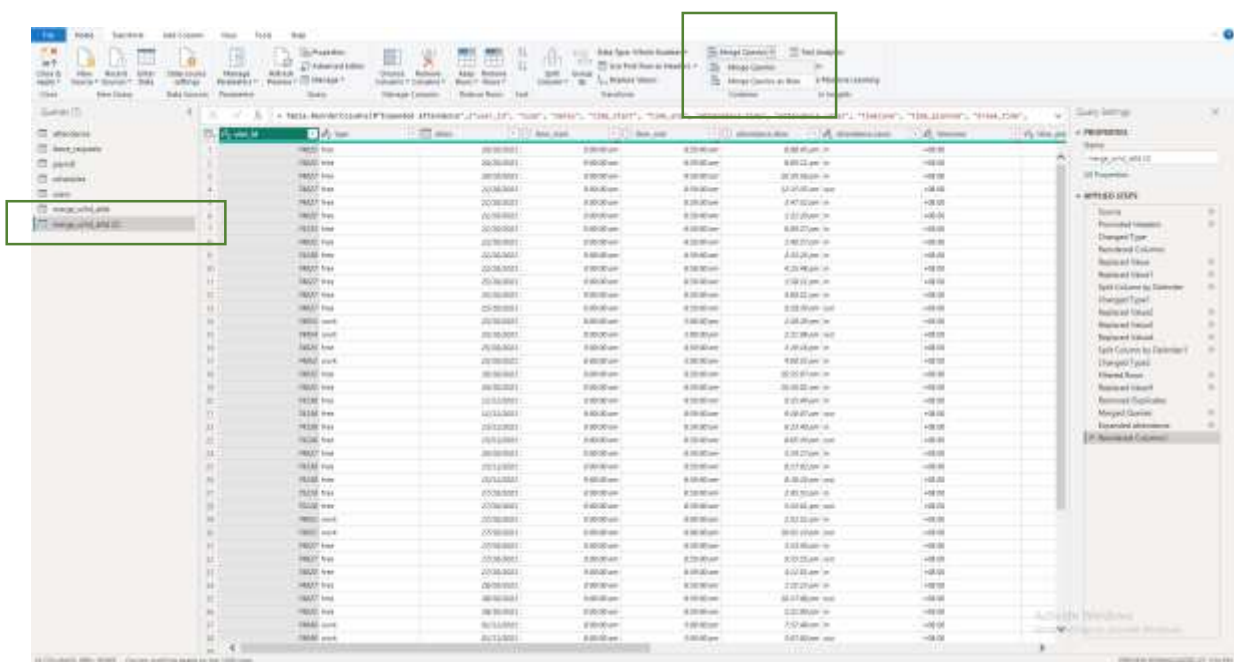
This is the result of moving to the position of attendance.time and attendance.cases

The screenshot shows the result of moving the 'attendance.cases' column. The columns are now arranged in pairs: attendance.time, attendance.cases, attendance.time, attendance.cases, attendance.time, attendance.cases. The data rows are visible, showing the time and cases for each record.

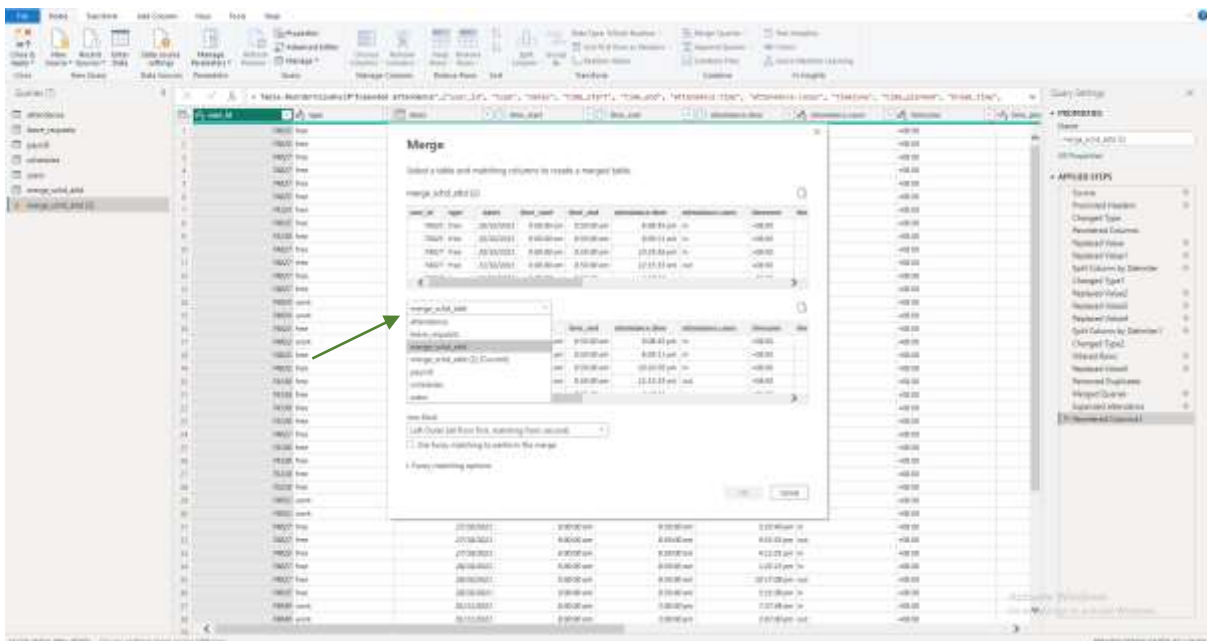
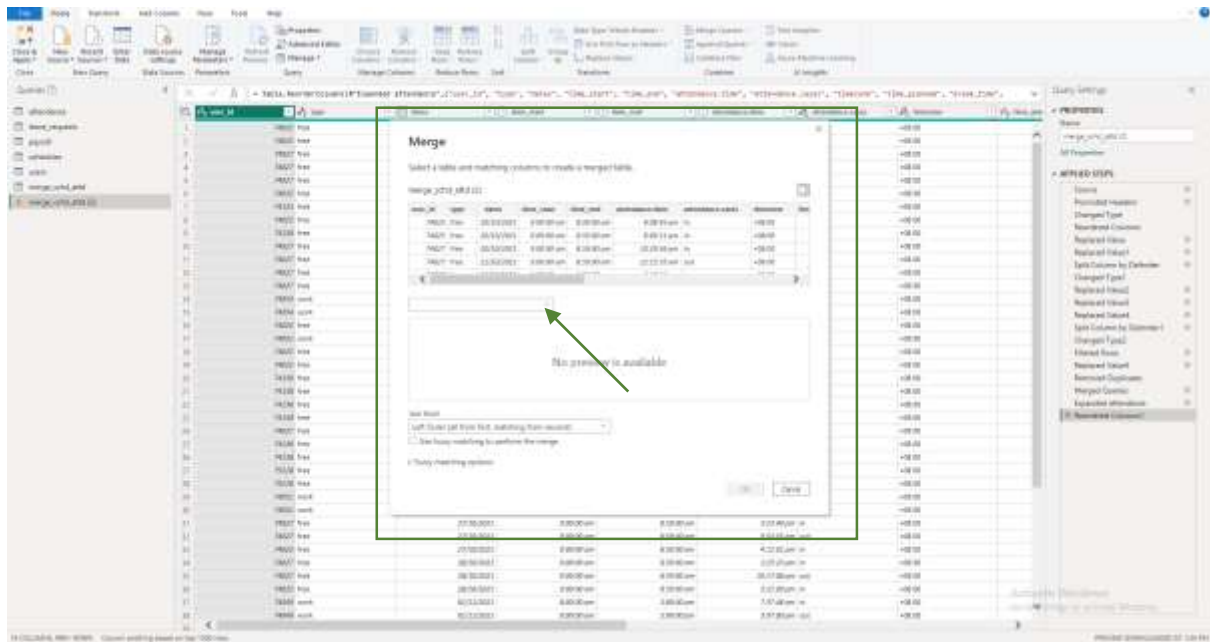
To arrange the data in one row, this is what we did,duplicate the merge_schd_attd table.



Now that we have duplicated the merge_schd_attd table, let's merge merge_schd_attd with merge_schd_attd (2). Go to Home, press the Merge Queries drop down arrow.



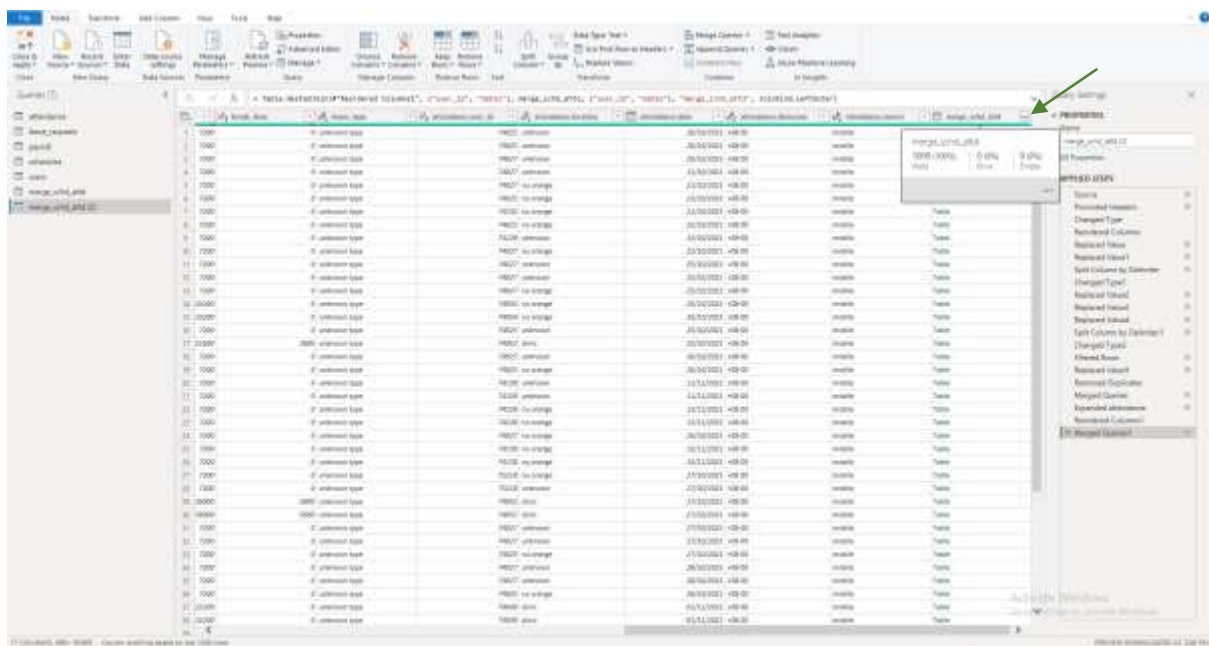
The Merge pop up window will appear. Press the drop down arrow to find the table named merge_schd_attd.



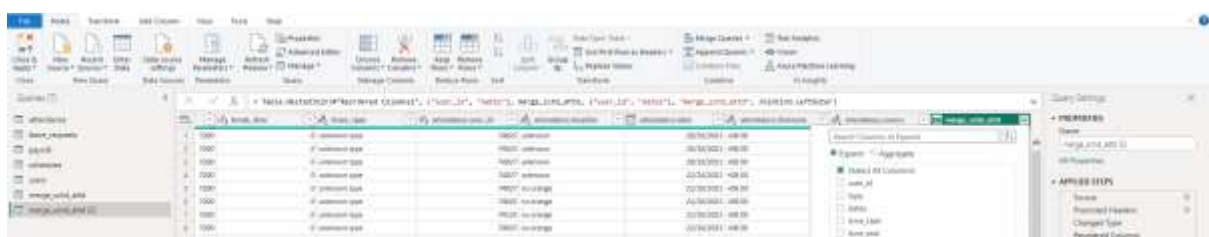
Set the commonality of the two tables using the user_id column and dates. Join Kind is Full Outer, and press ok.

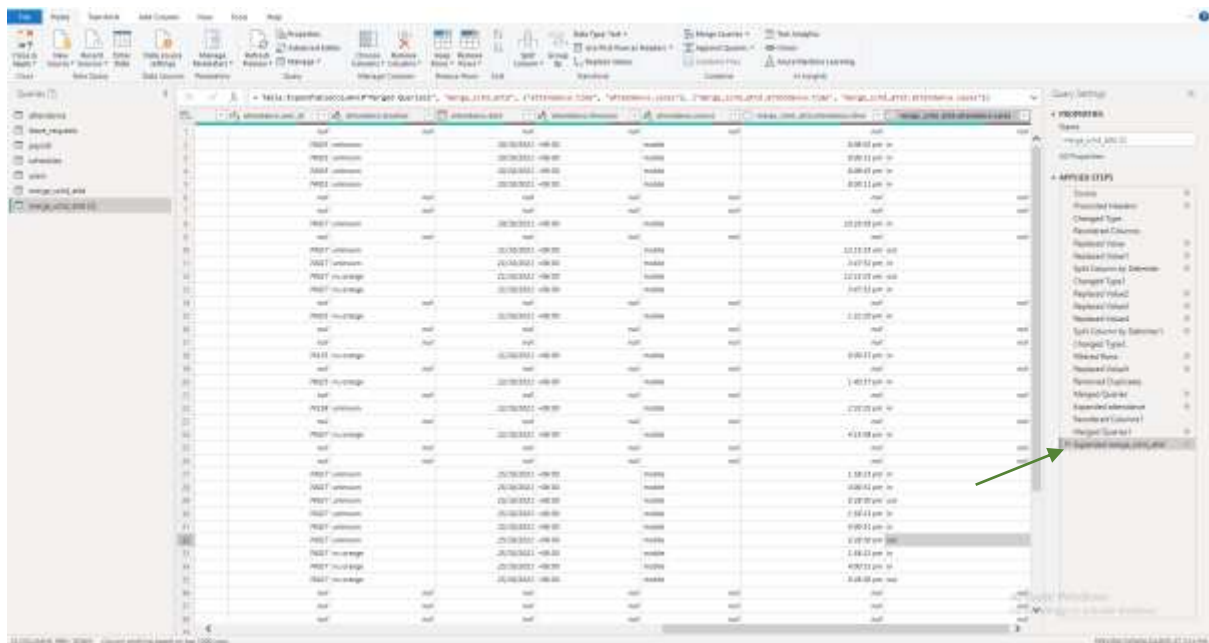


The new merge_schd_attd column will appear. We have to press the icon next to the column name to expand and show the result of the merging process that we did in the two tables.



Leave only the attendance.time and attendance.cases column checked and click ok.



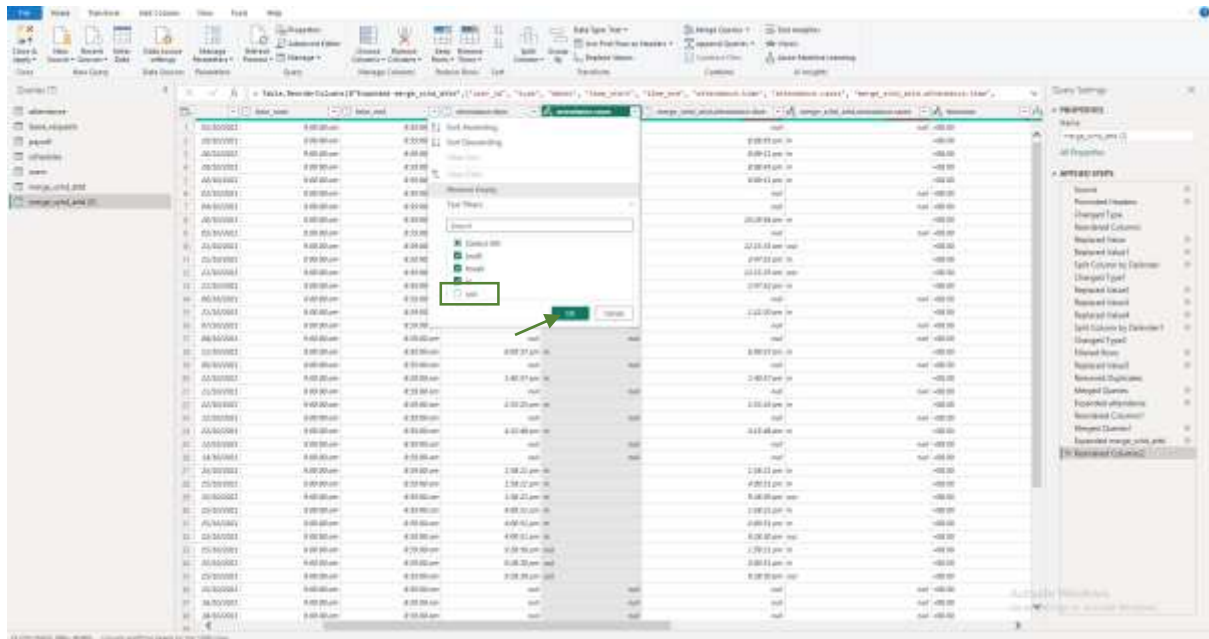


This is the result, the two new columns are next to attendance.cases.

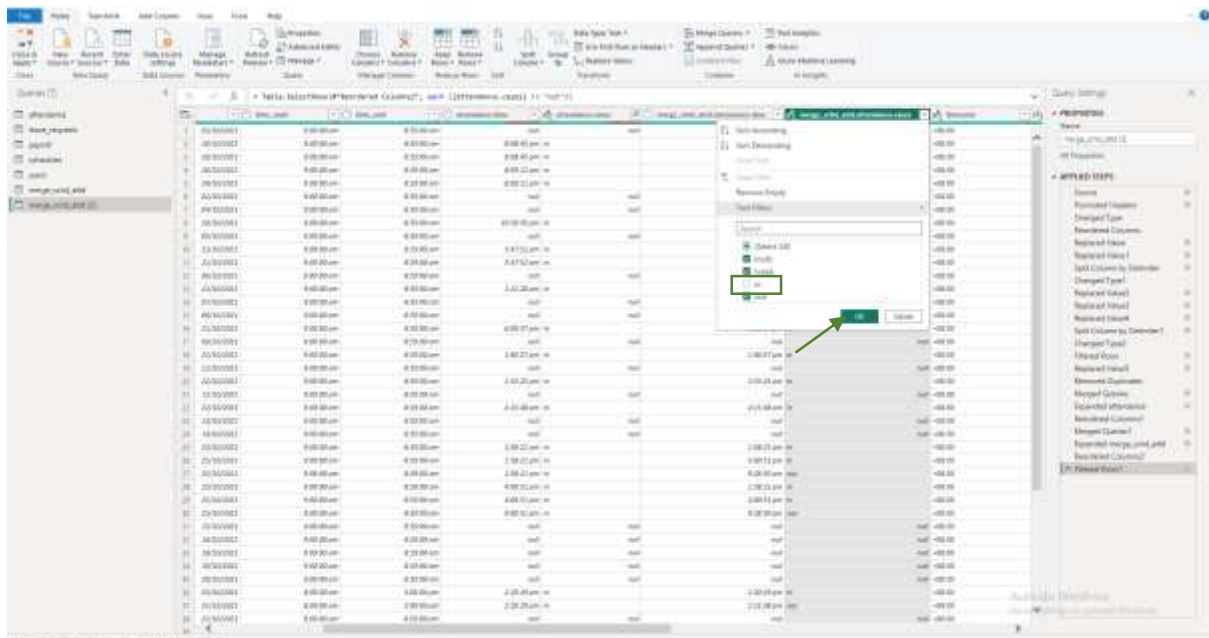
The screenshot displays the Microsoft Access interface with a query named 'qryAttendanceCases' open in Datasheet View. The query result grid shows data for various attendance cases, with two new columns added next to the 'attendance.cases' column. The new columns are labeled 'New Case' and 'New Case 2'. The data in the grid is as follows:

attendance.cases	New Case	New Case 2
1	10/10/2011	10/10/2011
2	10/10/2011	10/10/2011
3	10/10/2011	10/10/2011
4	10/10/2011	10/10/2011
5	10/10/2011	10/10/2011
6	10/10/2011	10/10/2011
7	10/10/2011	10/10/2011
8	10/10/2011	10/10/2011
9	10/10/2011	10/10/2011
10	10/10/2011	10/10/2011
11	10/10/2011	10/10/2011
12	10/10/2011	10/10/2011
13	10/10/2011	10/10/2011
14	10/10/2011	10/10/2011
15	10/10/2011	10/10/2011
16	10/10/2011	10/10/2011
17	10/10/2011	10/10/2011
18	10/10/2011	10/10/2011
19	10/10/2011	10/10/2011
20	10/10/2011	10/10/2011
21	10/10/2011	10/10/2011
22	10/10/2011	10/10/2011
23	10/10/2011	10/10/2011
24	10/10/2011	10/10/2011
25	10/10/2011	10/10/2011
26	10/10/2011	10/10/2011
27	10/10/2011	10/10/2011
28	10/10/2011	10/10/2011
29	10/10/2011	10/10/2011
30	10/10/2011	10/10/2011
31	10/10/2011	10/10/2011
32	10/10/2011	10/10/2011
33	10/10/2011	10/10/2011
34	10/10/2011	10/10/2011
35	10/10/2011	10/10/2011
36	10/10/2011	10/10/2011
37	10/10/2011	10/10/2011
38	10/10/2011	10/10/2011
39	10/10/2011	10/10/2011
40	10/10/2011	10/10/2011
41	10/10/2011	10/10/2011
42	10/10/2011	10/10/2011
43	10/10/2011	10/10/2011
44	10/10/2011	10/10/2011
45	10/10/2011	10/10/2011
46	10/10/2011	10/10/2011
47	10/10/2011	10/10/2011
48	10/10/2011	10/10/2011
49	10/10/2011	10/10/2011
50	10/10/2011	10/10/2011
51	10/10/2011	10/10/2011
52	10/10/2011	10/10/2011
53	10/10/2011	10/10/2011
54	10/10/2011	10/10/2011
55	10/10/2011	10/10/2011
56	10/10/2011	10/10/2011
57	10/10/2011	10/10/2011
58	10/10/2011	10/10/2011
59	10/10/2011	10/10/2011
60	10/10/2011	10/10/2011
61	10/10/2011	10/10/2011
62	10/10/2011	10/10/2011
63	10/10/2011	10/10/2011
64	10/10/2011	10/10/2011
65	10/10/2011	10/10/2011
66	10/10/2011	10/10/2011
67	10/10/2011	10/10/2011
68	10/10/2011	10/10/2011
69	10/10/2011	10/10/2011
70	10/10/2011	10/10/2011
71	10/10/2011	10/10/2011
72	10/10/2011	10/10/2011
73	10/10/2011	10/10/2011
74	10/10/2011	10/10/2011
75	10/10/2011	10/10/2011
76	10/10/2011	10/10/2011
77	10/10/2011	10/10/2011
78	10/10/2011	10/10/2011
79	10/10/2011	10/10/2011
80	10/10/2011	10/10/2011
81	10/10/2011	10/10/2011
82	10/10/2011	10/10/2011
83	10/10/2011	10/10/2011
84	10/10/2011	10/10/2011
85	10/10/2011	10/10/2011
86	10/10/2011	10/10/2011
87	10/10/2011	10/10/2011
88	10/10/2011	10/10/2011
89	10/10/2011	10/10/2011
90	10/10/2011	10/10/2011
91	10/10/2011	10/10/2011
92	10/10/2011	10/10/2011
93	10/10/2011	10/10/2011
94	10/10/2011	10/10/2011
95	10/10/2011	10/10/2011
96	10/10/2011	10/10/2011
97	10/10/2011	10/10/2011
98	10/10/2011	10/10/2011
99	10/10/2011	10/10/2011
100	10/10/2011	10/10/2011

In attendance.cases press the drop down arrow next to the column name. uncheck the “out” value and then click ok.



In merge_schd_attd.attendance.cases press the drop down arrow next to the column name. uncheck the “in” value and then click ok.



The screenshot displays the SAP S/4HANA Fiori 'Manage Material' app. The main table lists materials with columns for Material, Description, Unit of Measure, and others. Two green arrows point from the left sidebar to the 'Material' column. A green box highlights the 'Material' column header and the first few rows of the material list. The right sidebar shows the 'Attributes' tab with various material attributes.

The total rows of our final table are 21247.

The screenshot shows a data table with the following columns: `date`, `time`, `attendance.time`, and `attendance.status`. The table contains 21,247 rows. A green arrow points to the `attendance.time` column header, indicating it is the target for a renaming operation.

Let's change the column name `attendance.time` to "login". To do this, double click on the column name to highlight it, then type "login" this will be the new column name

The screenshot shows the same data table after the column name change. The column `attendance.time` has been renamed to `login`. The table structure remains the same, but the column name is updated.

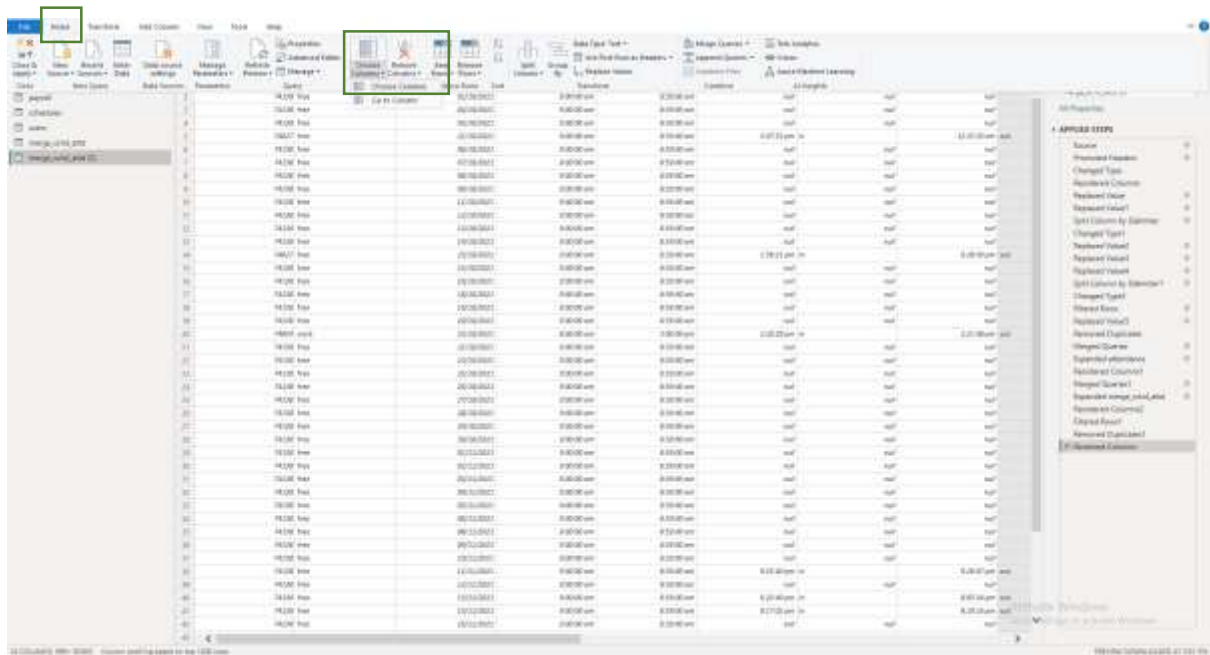
The previous attendance.time column name has become "login". Repeat this process to replace the attendance.cases column with "cases-in", merge_schd_attd.attendance.time with "log out", merge_schd_attd.attendance.cases with "case-out". This is to make it look better and easier to understand.

The screenshot shows a data table with the following columns: date, time, login, cases-in, log out, case-out, and several other columns. The 'login' column is highlighted with a green box. Green arrows point to the 'cases-in', 'log out', and 'case-out' columns. The table contains data for various dates and times, with some rows showing 'login' and 'log out' times.

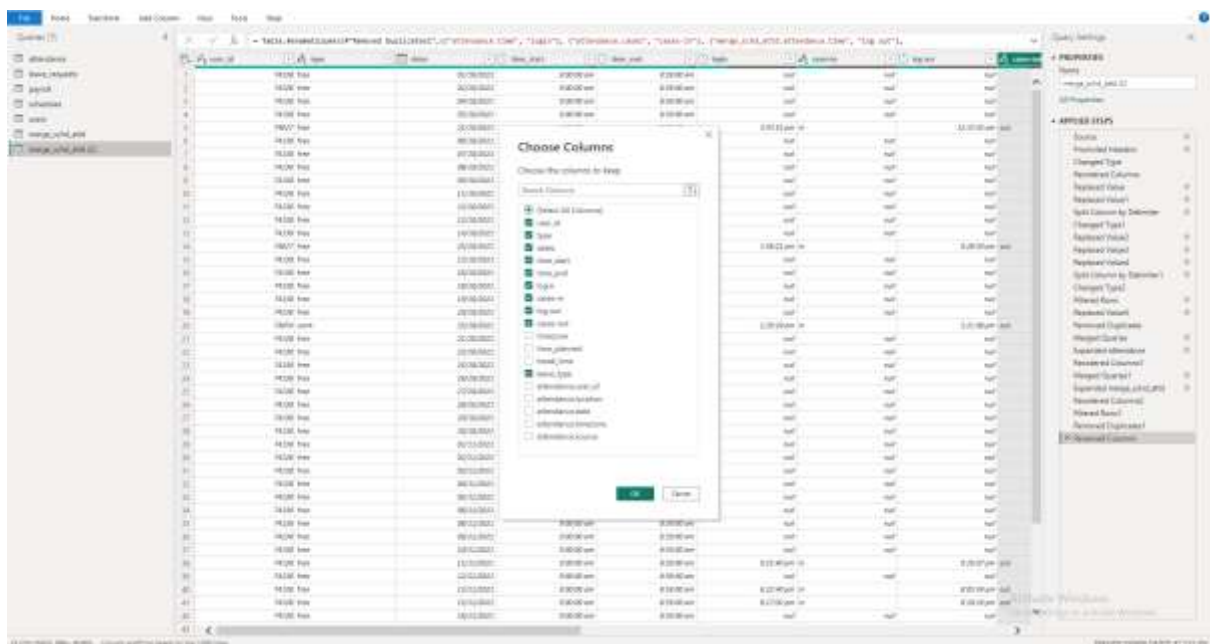
We have changed the column name of four columns.

The screenshot shows the same data table as before, but with the column names updated. The 'login' column is highlighted with a green box. A green arrow points to the 'case-out' column. The table contains data for various dates and times, with some rows showing 'login' and 'log out' times.

Let's delete other columns that are not needed in our table. Go to Home, press the drop down arrow of Choose Column and click Choose Column.



Leave the checkbox of user_id,type,dates,time_start,time_end,login,cases-in,log out,cases-out and leave_type checked and uncheck the remaining column to remove it from this table. Click ok to continue.



This is the final result of data cleaning and transformation of the dataset. There are 10 columns with 21247 rows.

The screenshot shows a data table with the following columns: `year_id`, `age`, `date`, `time_point`, `time_end`, `height`, `weight`, `leg_len`, `name`, and `sex`. The table contains 21,247 rows of data. On the right side, the 'APPLIED STEPS' panel is visible, showing a list of transformations. A green arrow points to the 'Removed Other Columns' button at the bottom of this panel.

Rename the final table as `merge_final_table`. Double click the table `merge_schd_attd (2)` to highlight and change the table name to `merge_final_table`.

This screenshot shows the same data table, but the 'merge_schd_attd (2)' table is highlighted in the left sidebar. The 'APPLIED STEPS' panel on the right is also visible, showing the same list of transformations. The table structure and data remain the same as in the previous screenshot.

After renaming the table, press the Close and Apply drop down arrow and click Close and Apply to load the dataset.

