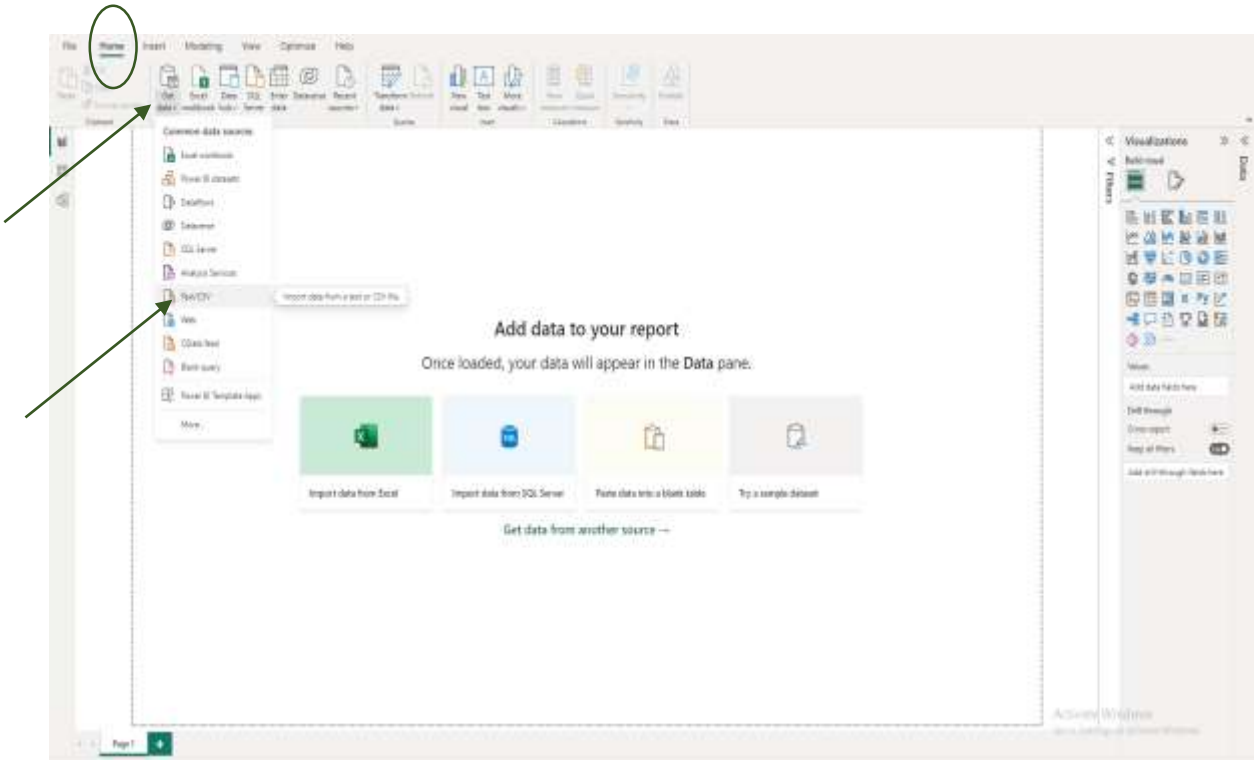
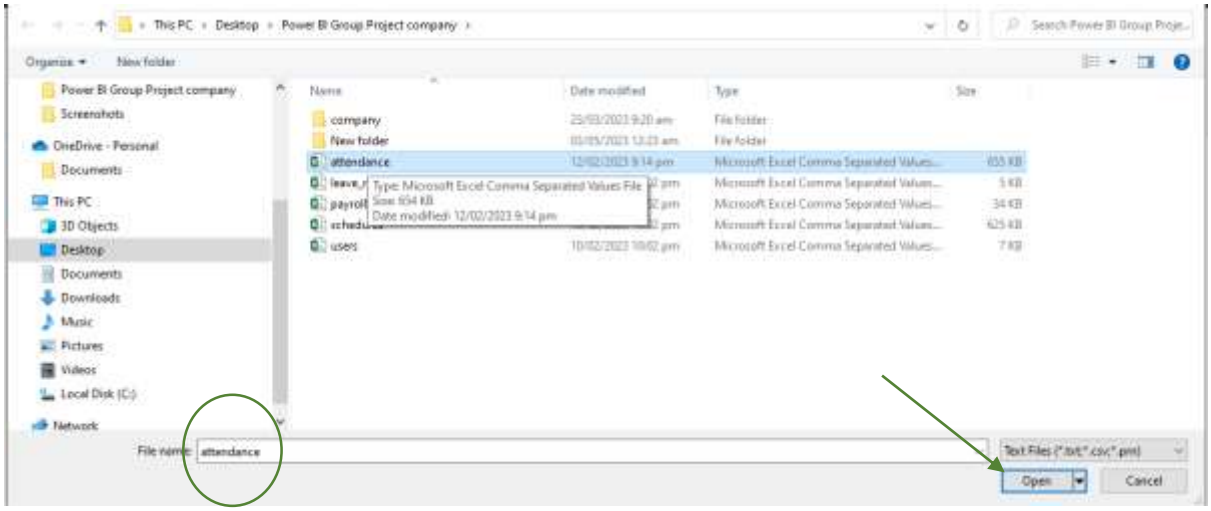
Methodology

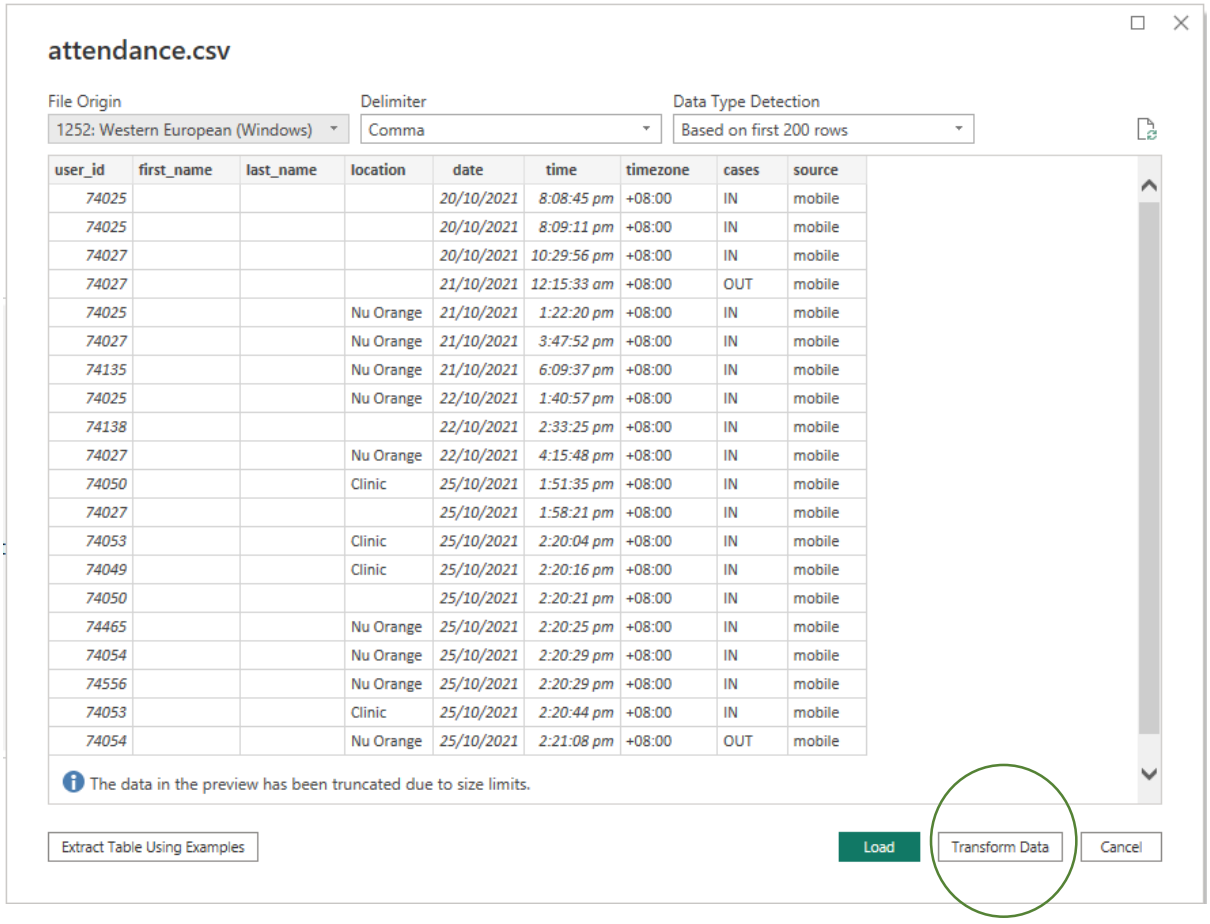
Importing Data to Power BI.Open Power BI Desktop on your PC. Go to Home and click Get Data then select Text/CSV.



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

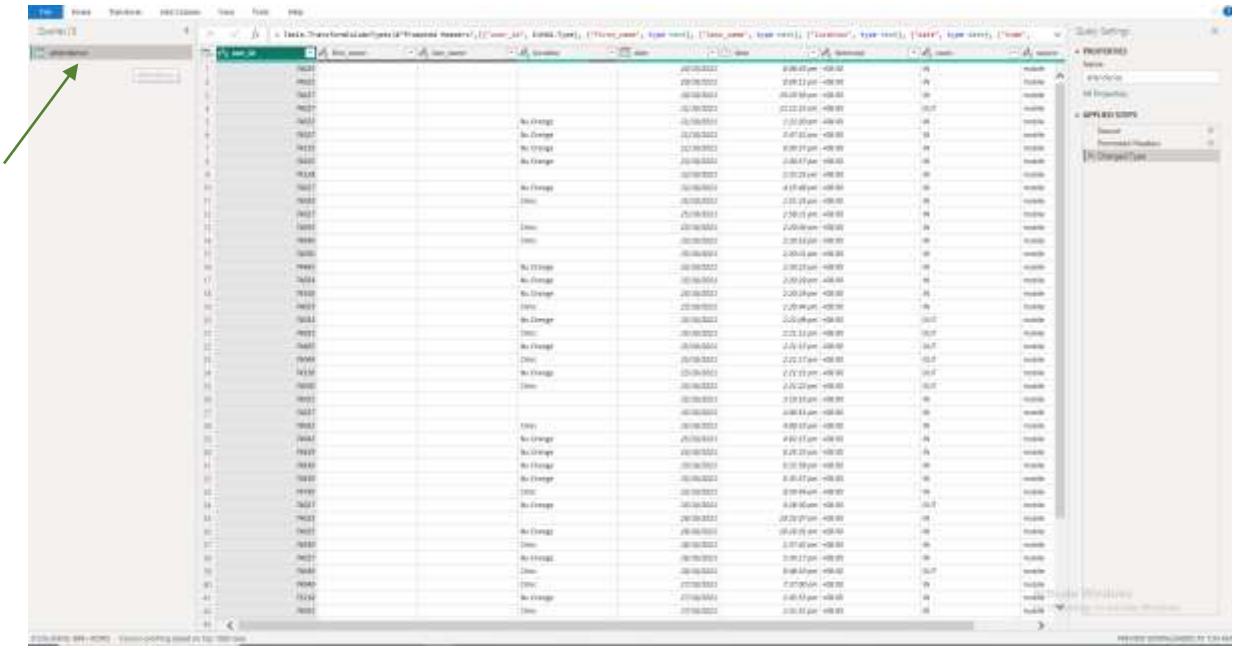


This new window will pop up to the screen. This window is the preview of the file you are about to import. Press Transform Data first, because we still need to clean and transform the data.

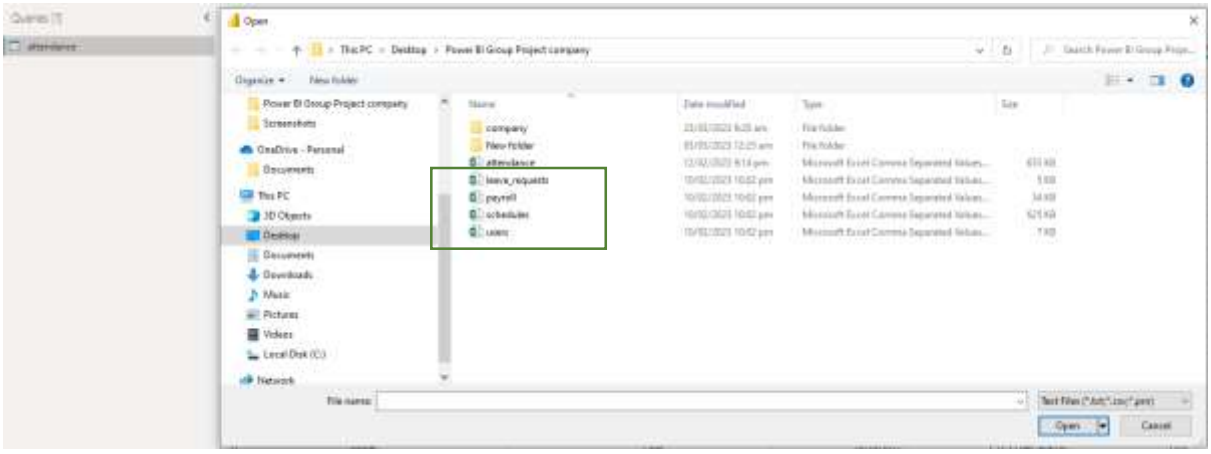


user_id	first_name	last_name	location	date	time	timezone	cases	source
74025				20/10/2021	8:08:45 pm	+08:00	IN	mobile
74025				20/10/2021	8:09:11 pm	+08:00	IN	mobile
74027				20/10/2021	10:29:56 pm	+08:00	IN	mobile
74027				21/10/2021	12:15:33 am	+08:00	OUT	mobile
74025			Nu Orange	21/10/2021	1:22:20 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

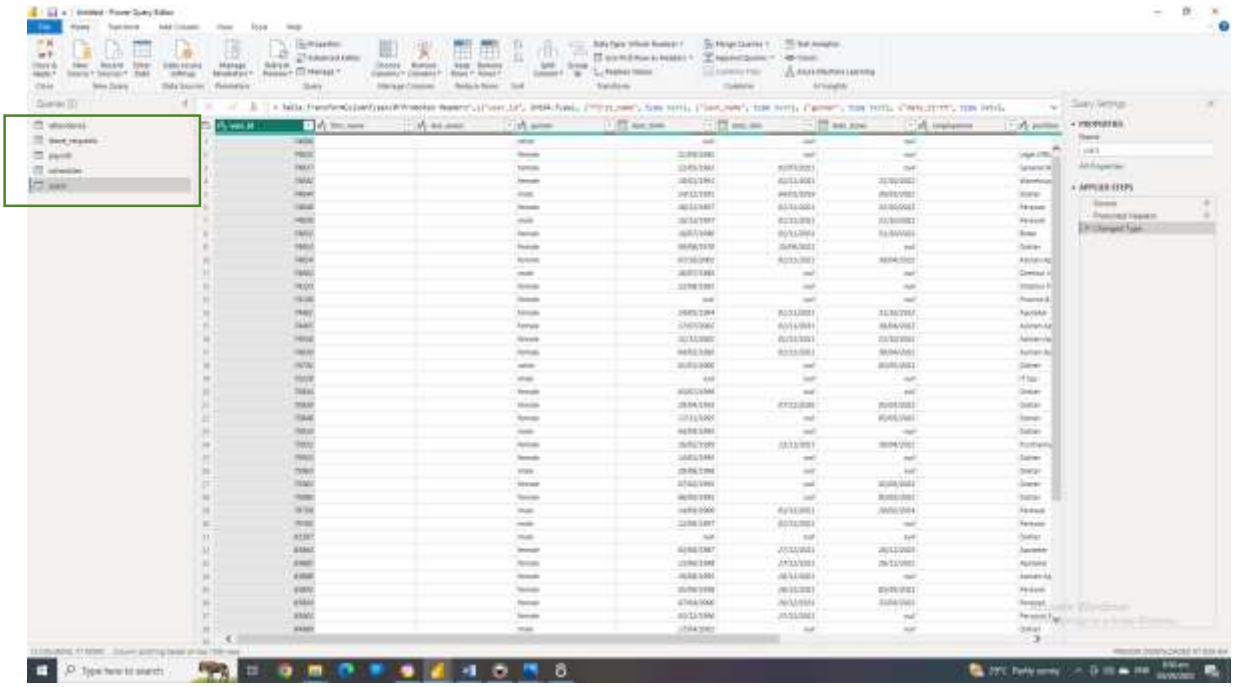
This is the final result of the steps we took to import the data.



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

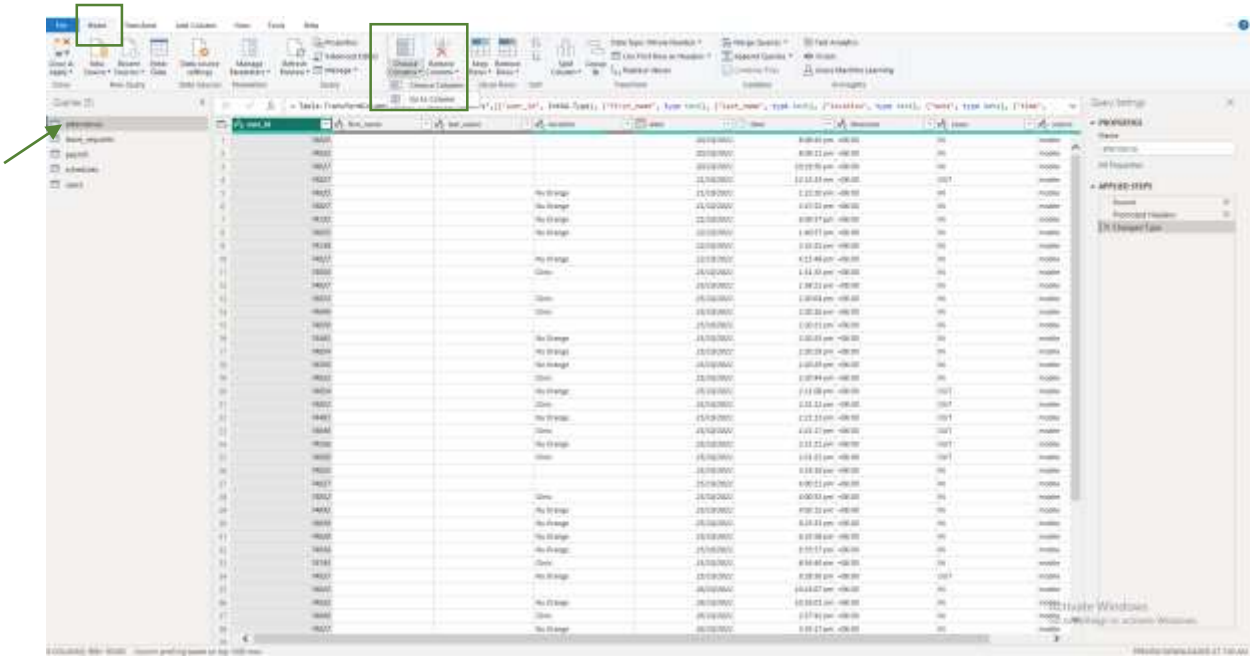


After all the files are imported into Power BI. The next step is the cleaning and transformation of the data.

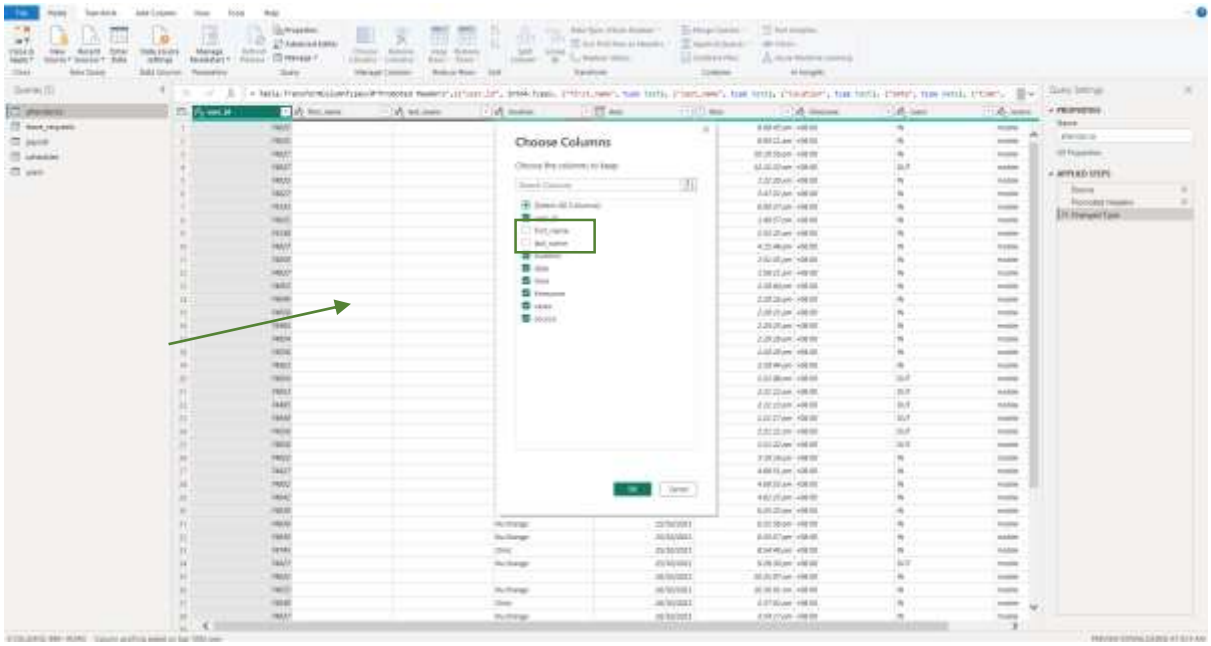


CLEANING AND TRANSFORMING ATTENDANCE TABLE

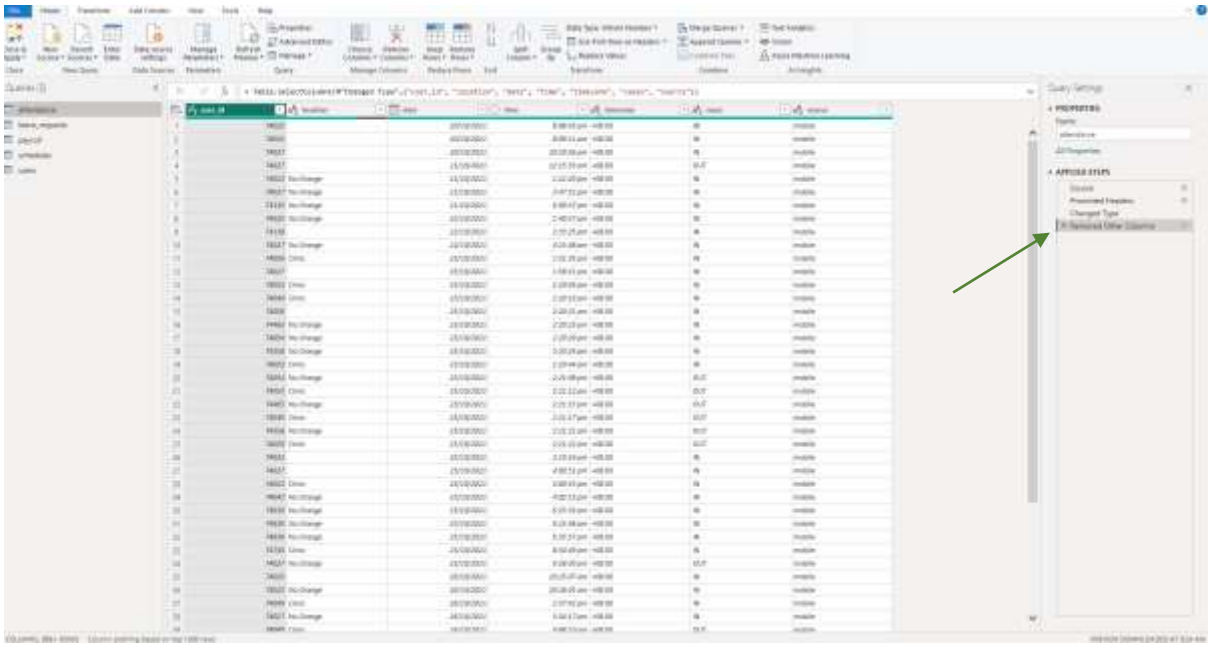
Removing unwanted columns attendance table. Select the attendance table in the Navigator Pane. Go to Home, click Choose Columns drop down arrow and click Choose Columns.



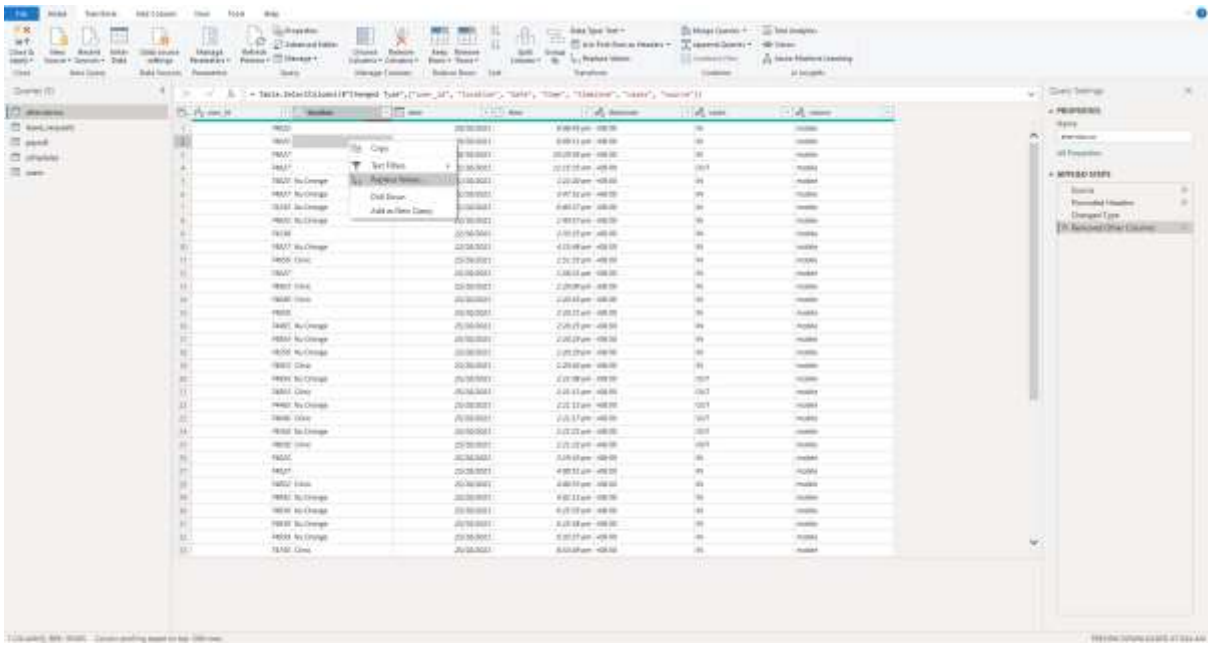
The Choose Columns pop up window appear. Unselect the check box that you want to remove from the attendance table. We chose to remove the first_name and last_name columns from the table because they have no data.



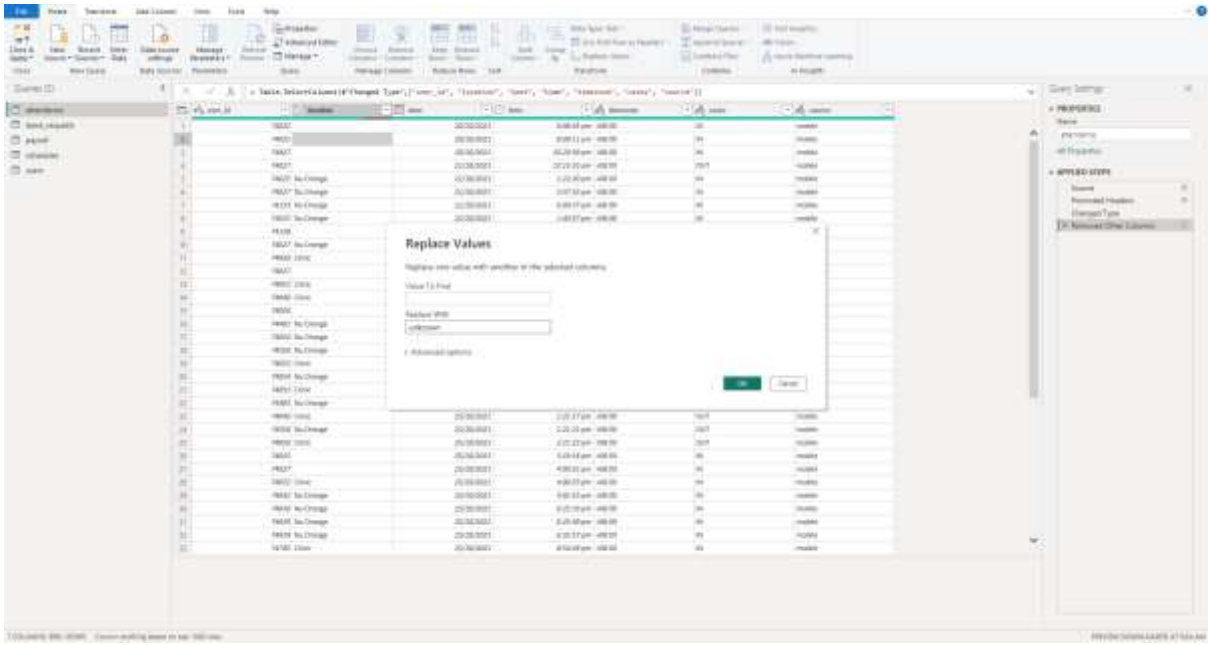
Removing unwanted columns result. See new applied steps.



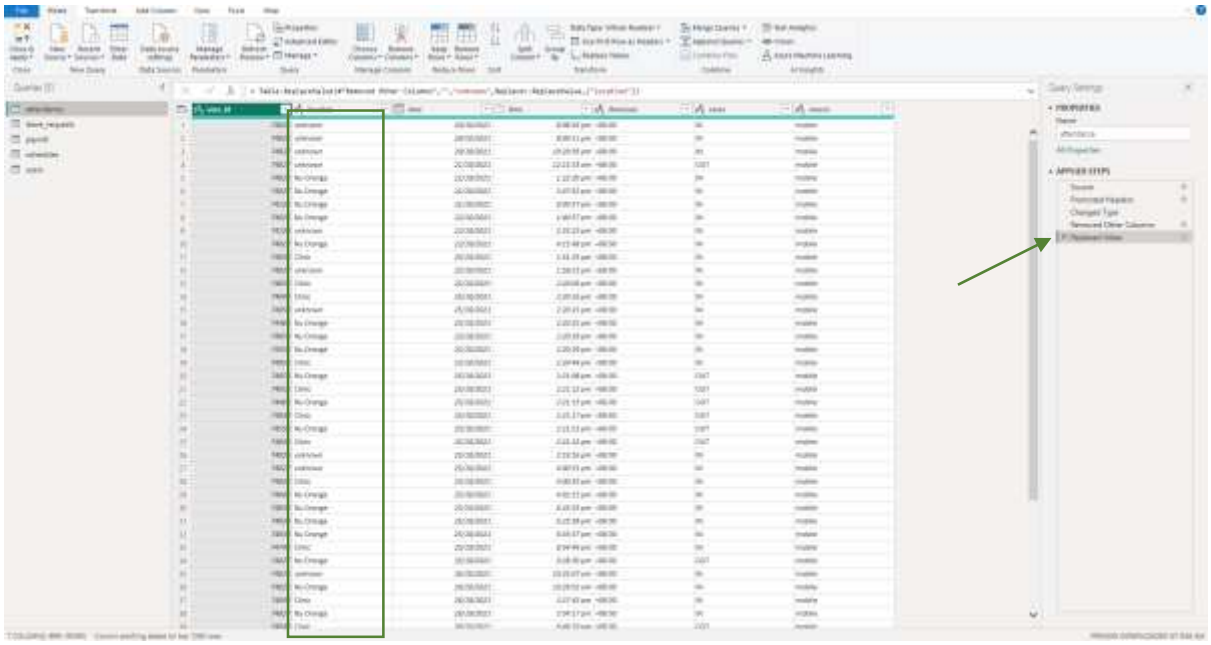
Replacing blank value of location column with “unknown”. Right click the cell of location column and select replace value.



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

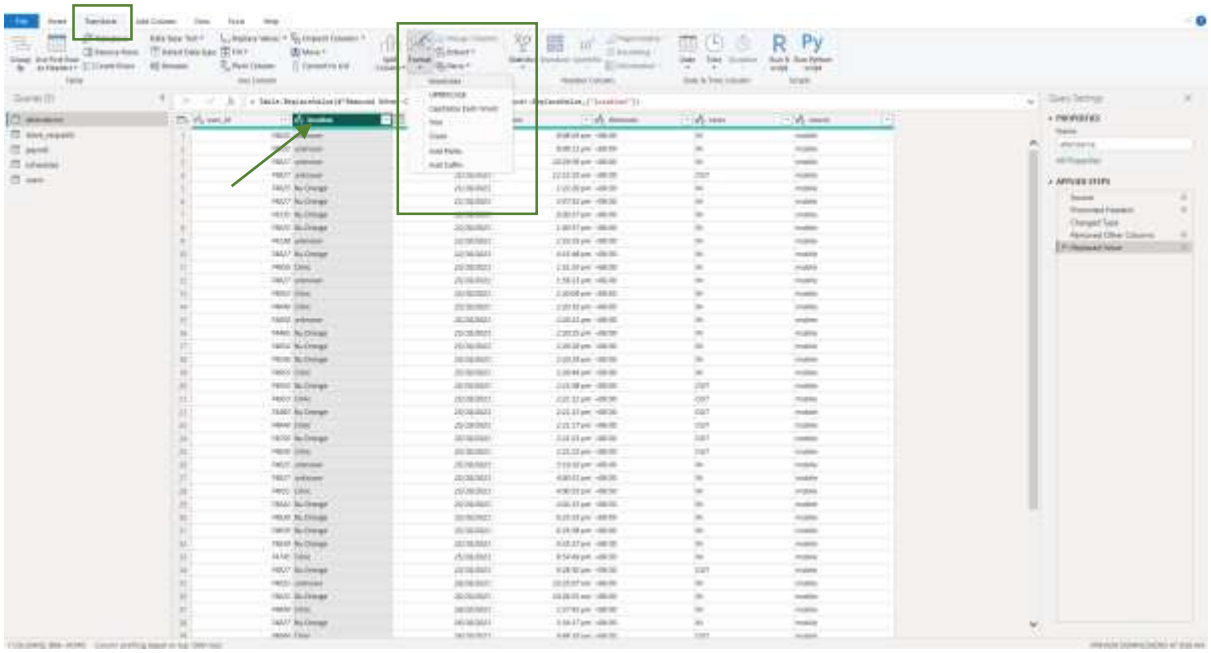


Result of replacing the blank value with unknown in location column.

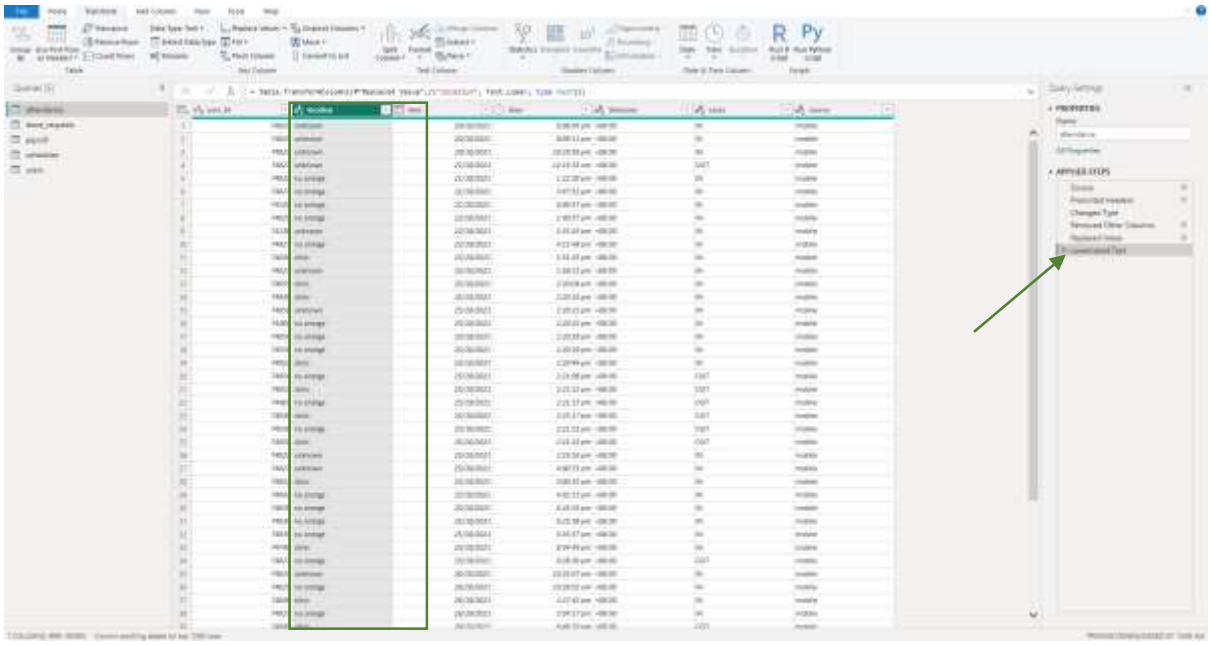


Changing the text format of the location column to lowercase letters.

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

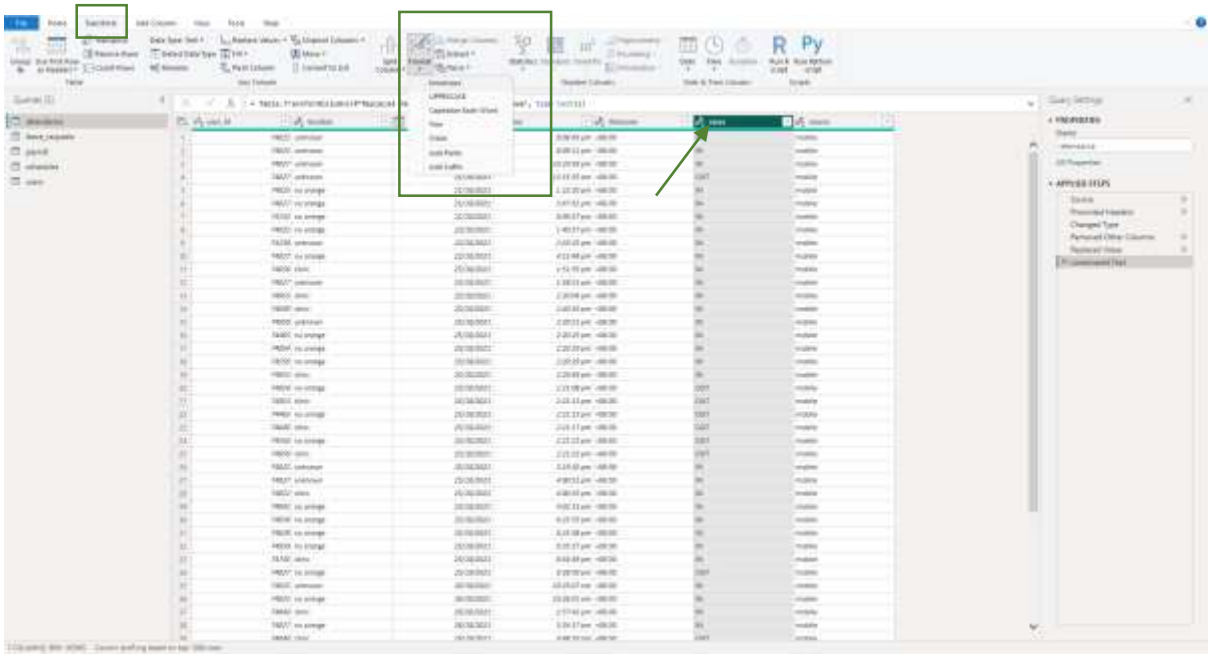


Result of location column formatted to lowercase letter.

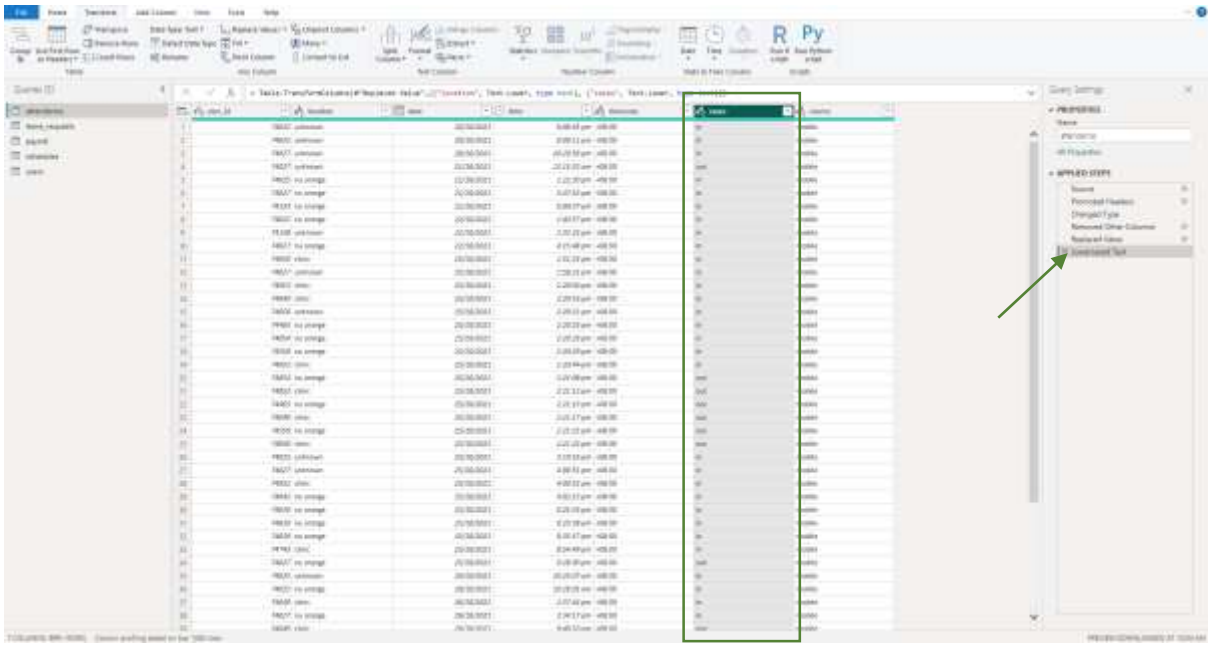


Changing the text format of the cases column to lowercase letters.

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

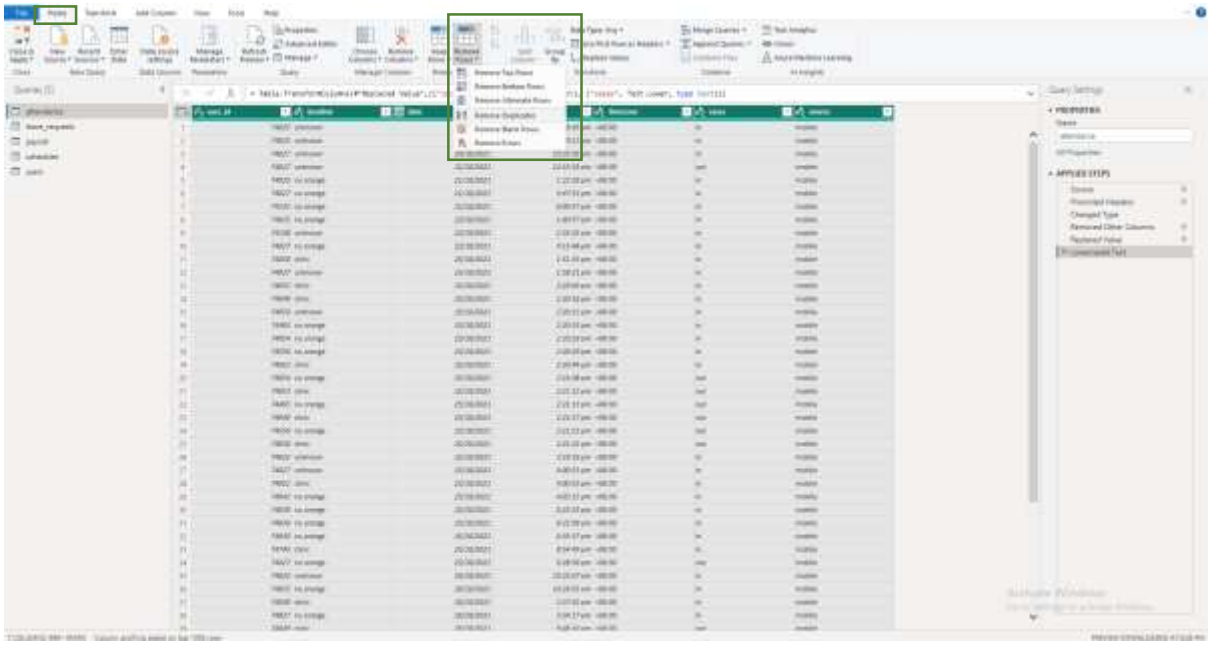


Result of cases column formatted to lowercase letters.

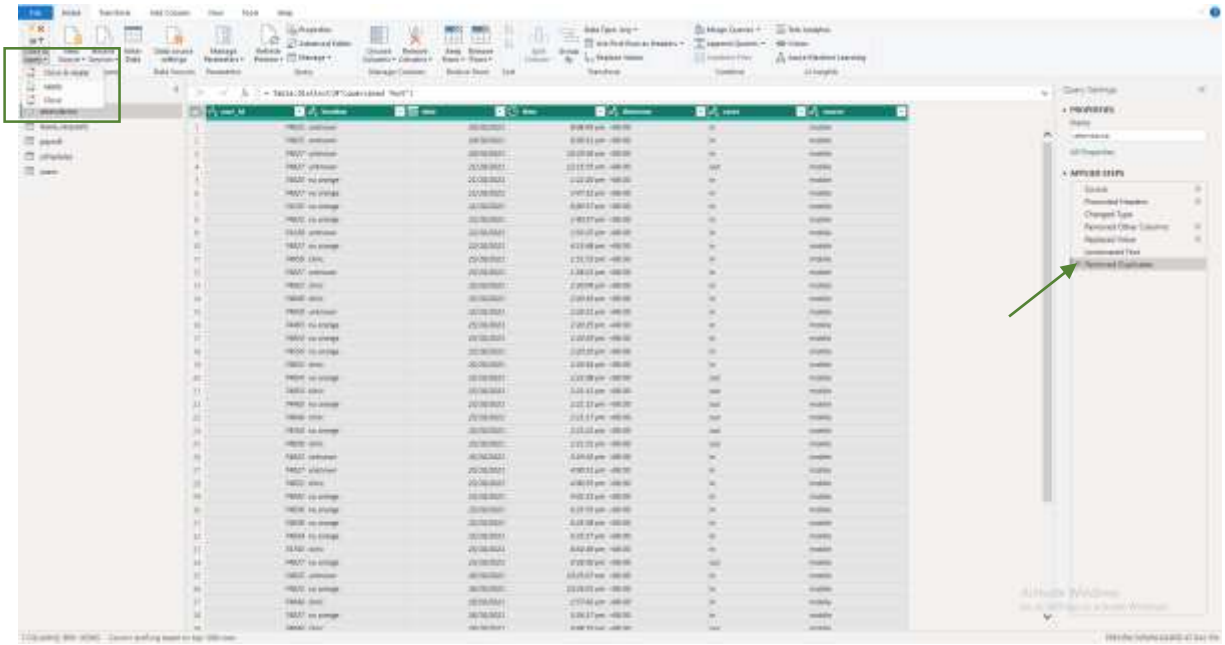


Removing duplicate of attendance table.

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

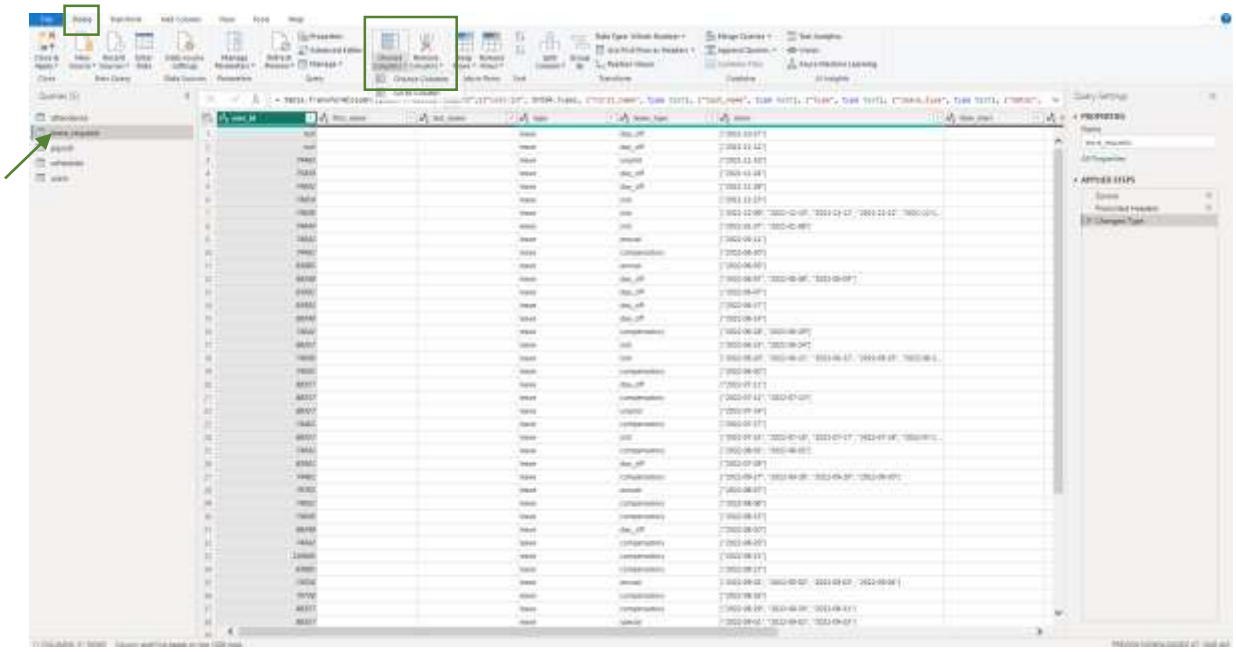


Result of Removing Duplicates of attendance table with 12604 rows.Click close and apply to load the data to model.

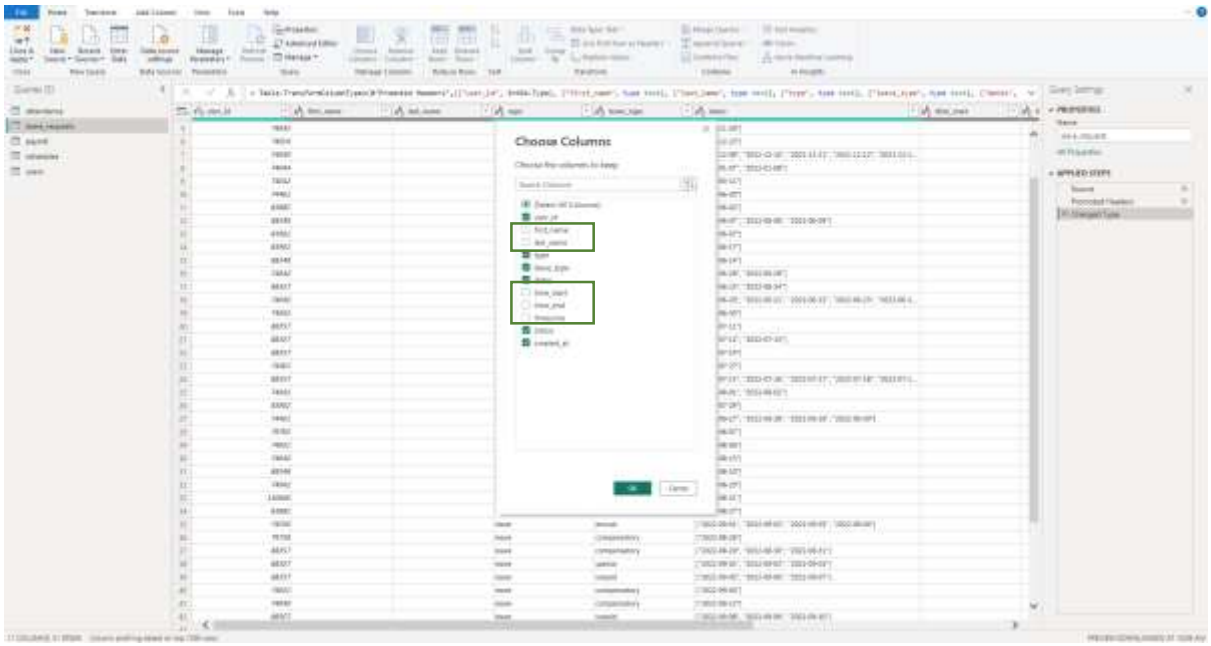


CLEANING AND TRANSFORMING LEAVE REQUEST TABLE

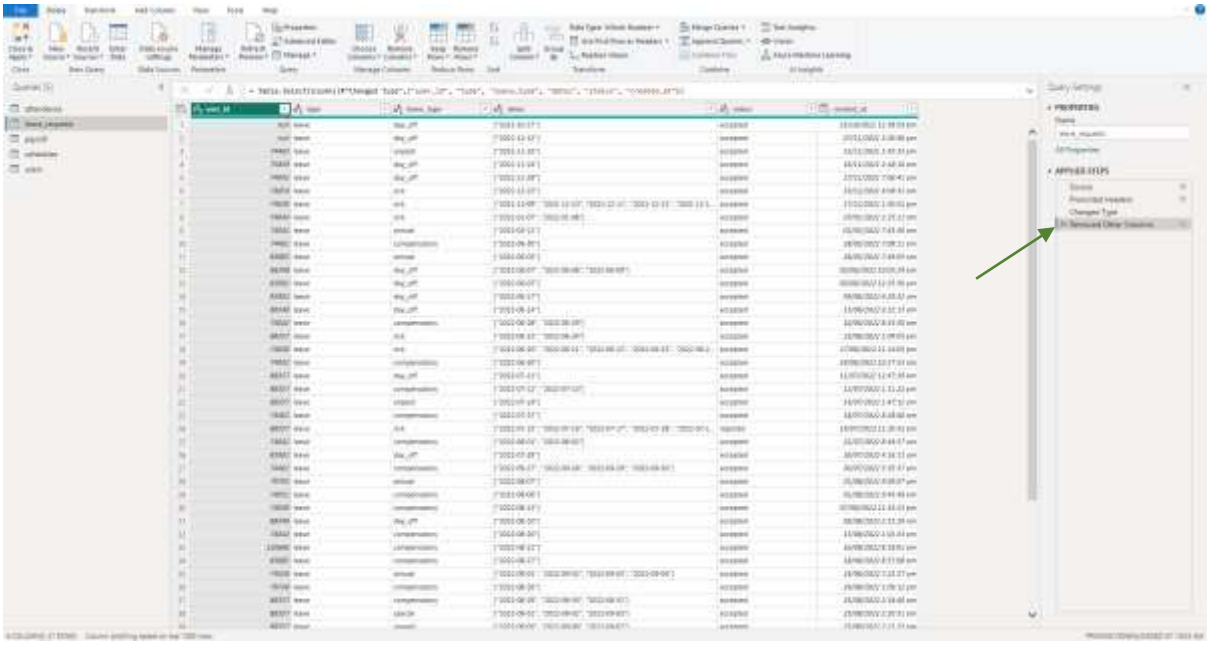
Removing unwanted columns of leave_requests table. Select the leave_requests table in the Navigator Pane. Go to Home, click Choose Columns dropdown arrow and then click Choose Columns.



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.

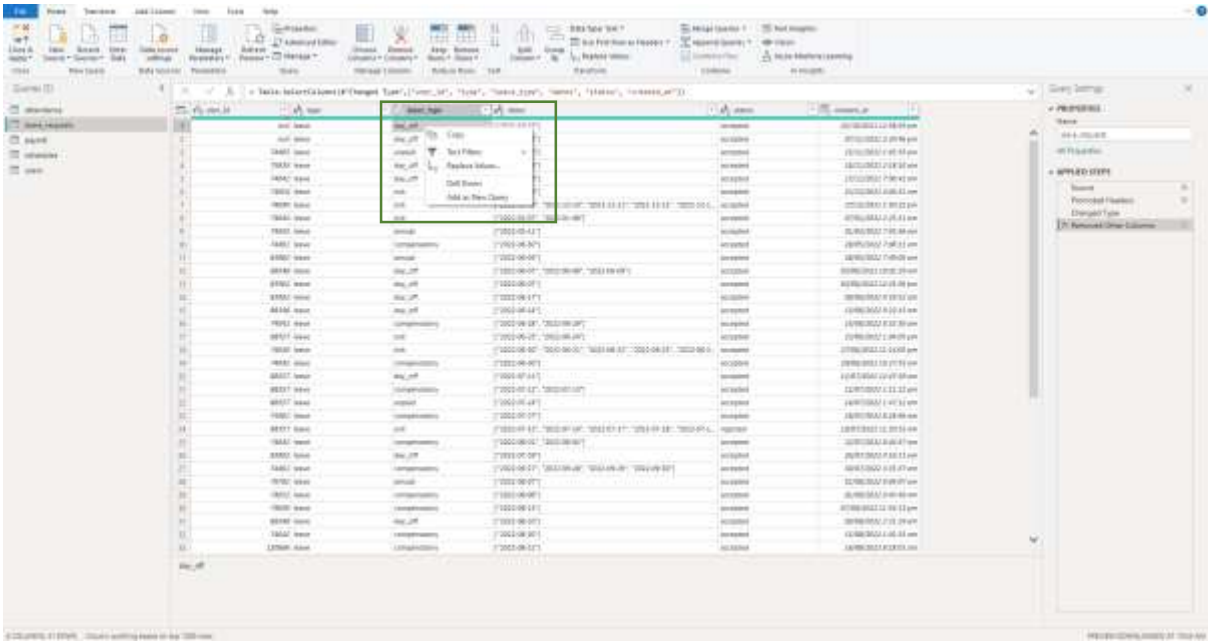


Result of removing unwanted column from leave_requests table.

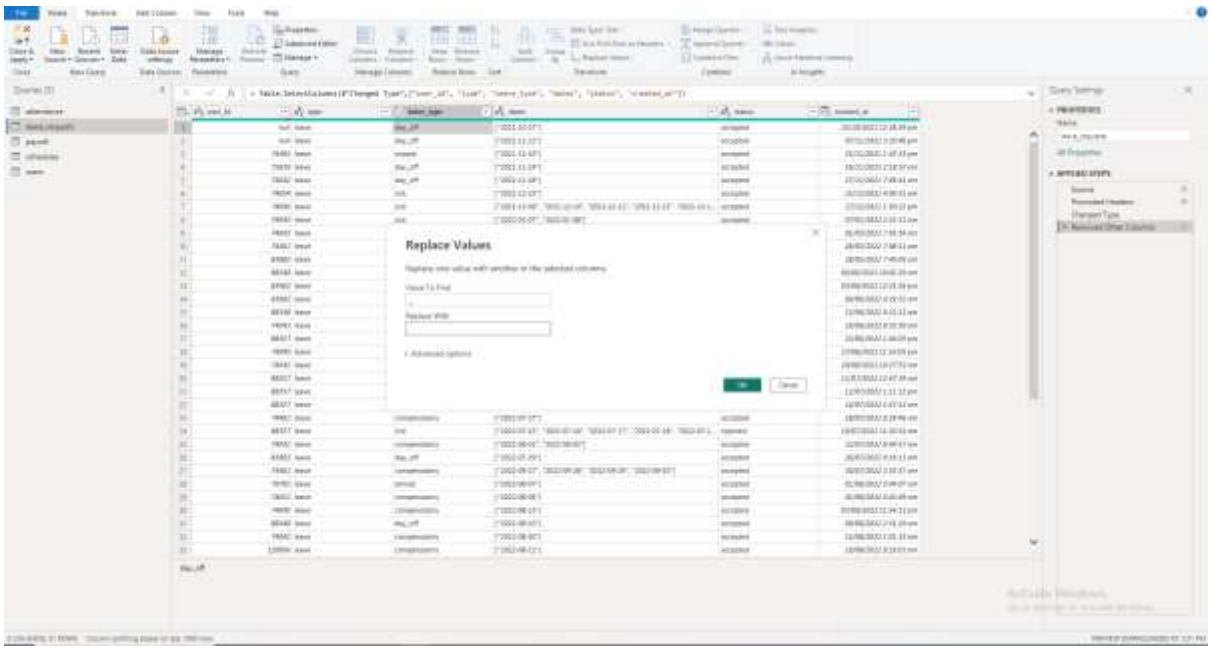


Replace the underscore of day_off with space in leave_type.

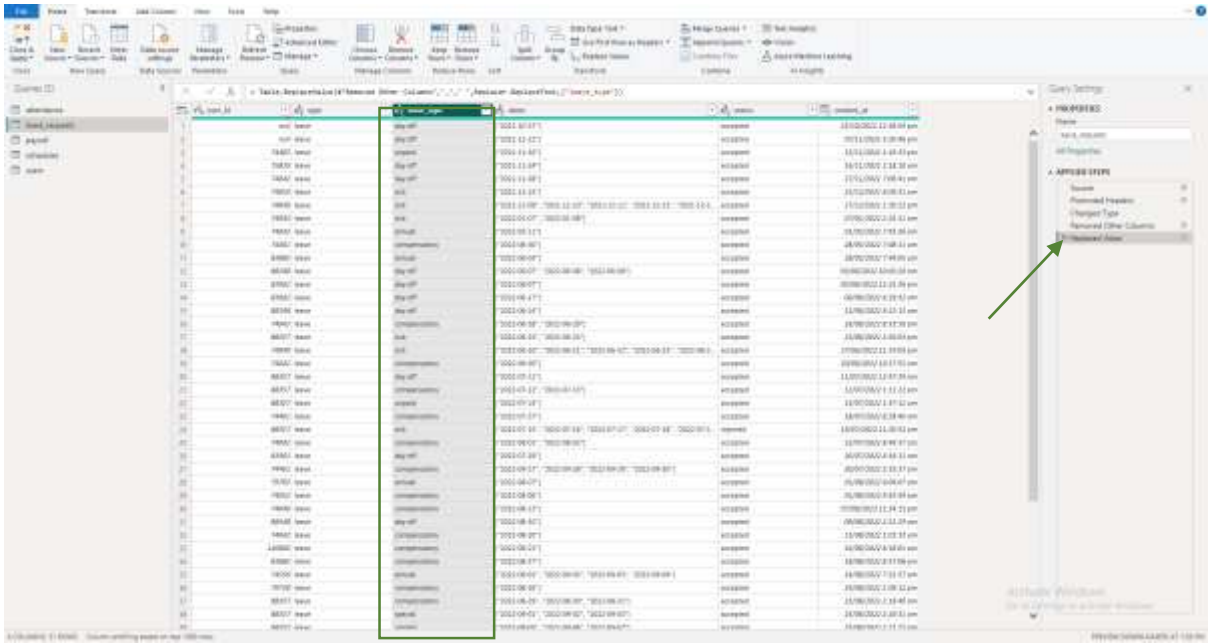
Right Click the cell of leave_type column which has a value of day_off, right click and select replace value.



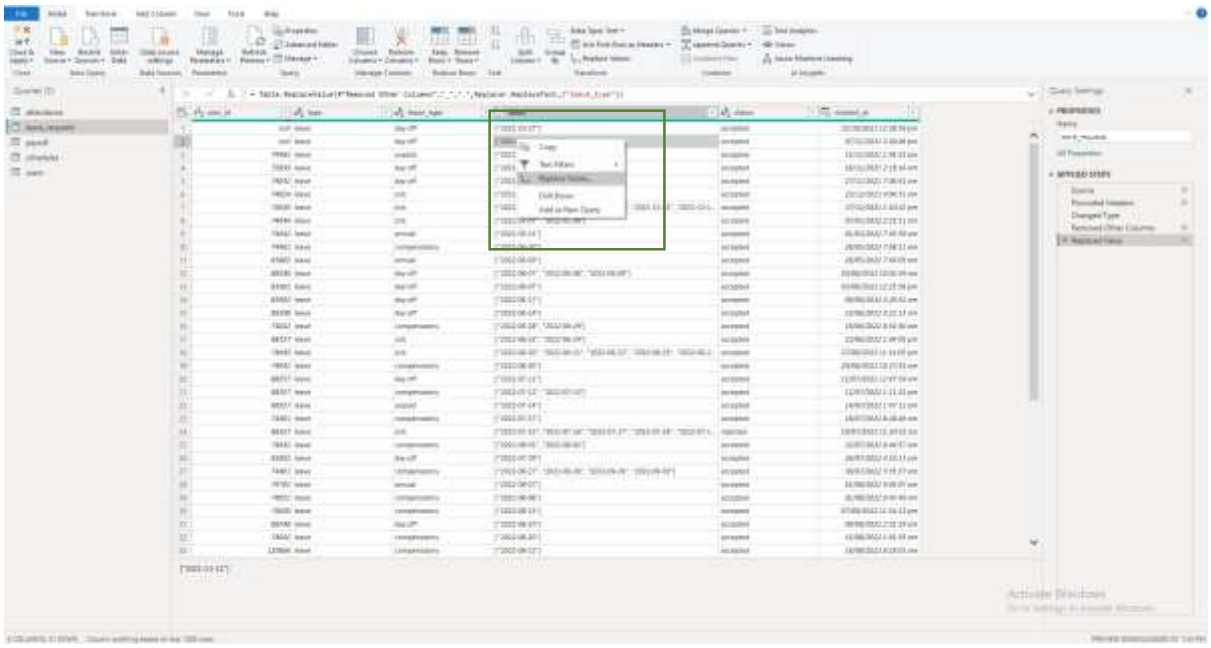
Input “_ “to Value to Find and inside the Replace With input box press the space bar and click ok.



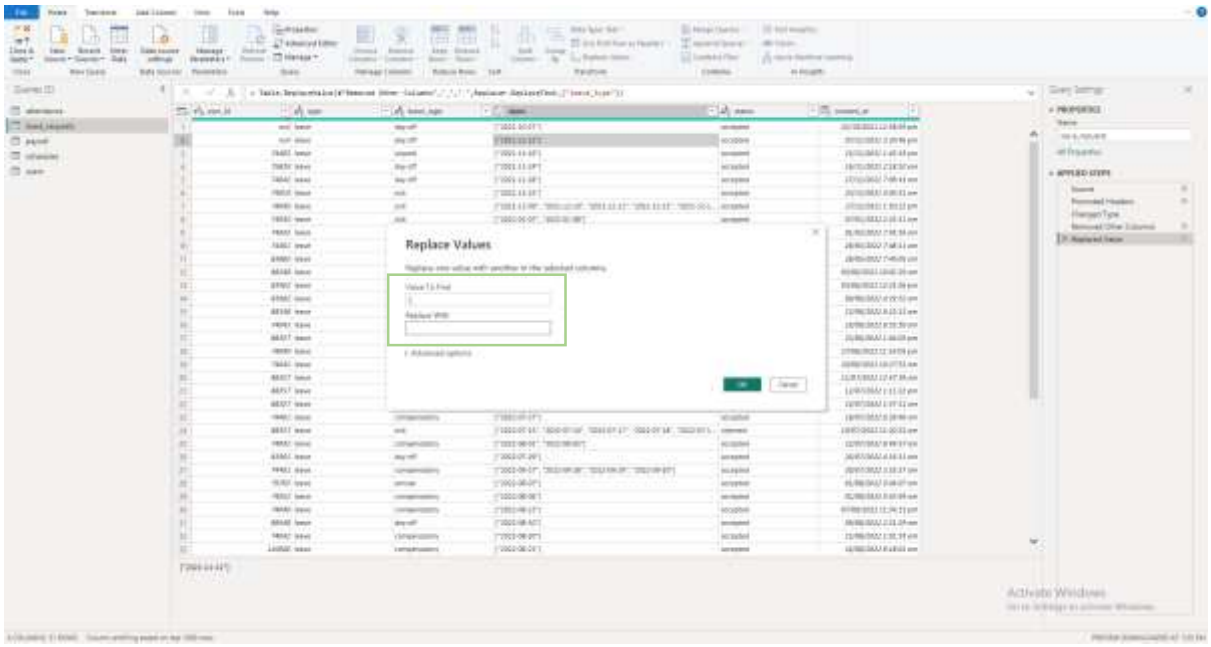
Result of replacing the day off underscore by space.



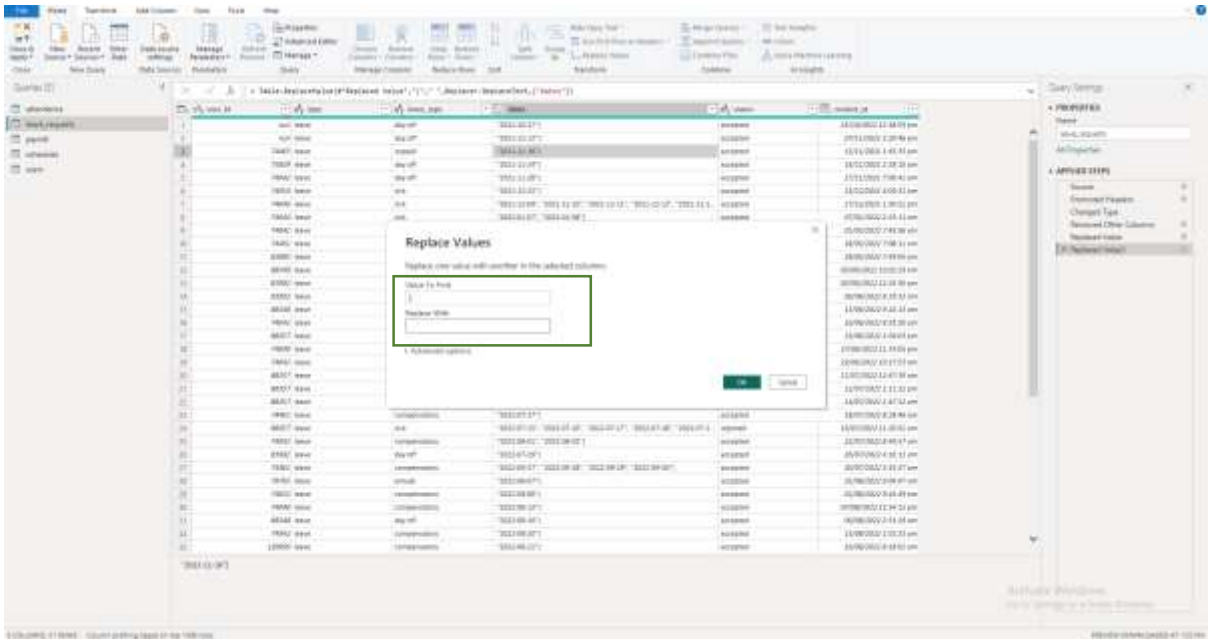
Cleaning the date column of the leave_requests table. Right click the cell of the dates column and click replace values.



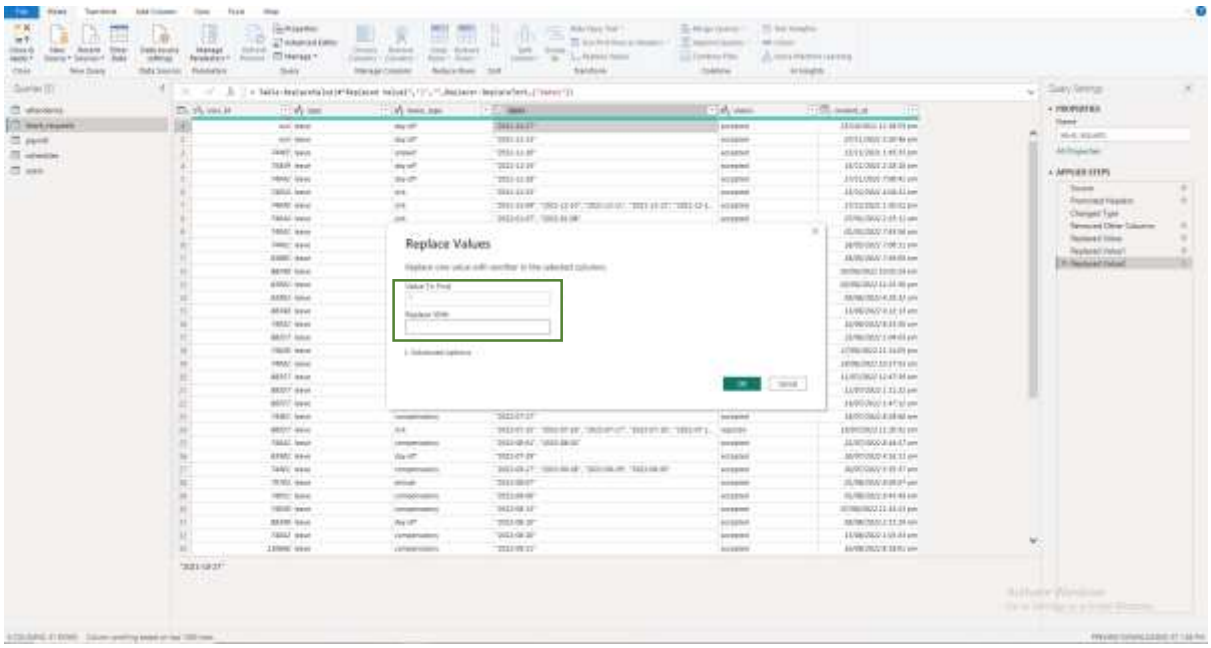
Input the open bracket to Values to Find and leave the Replace With input box blank and then click ok.



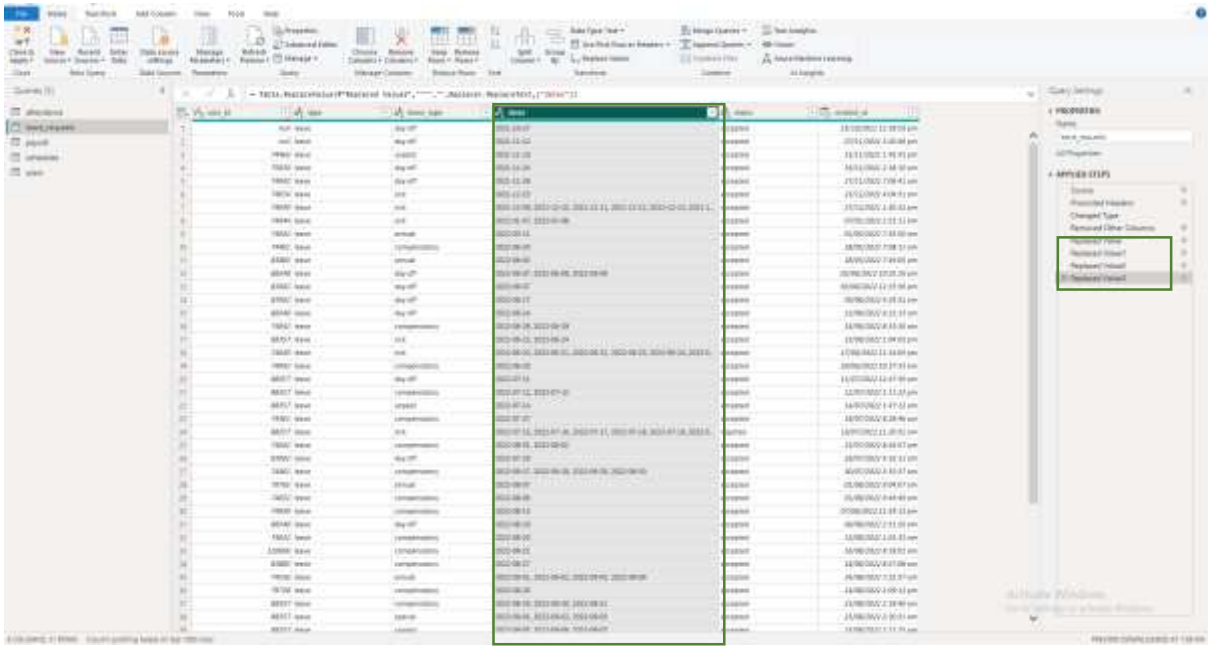
Right click again the cell of the dates column and select replace values. Input the close bracket to Values to Find and leave the Replace With input box blank and then click ok.



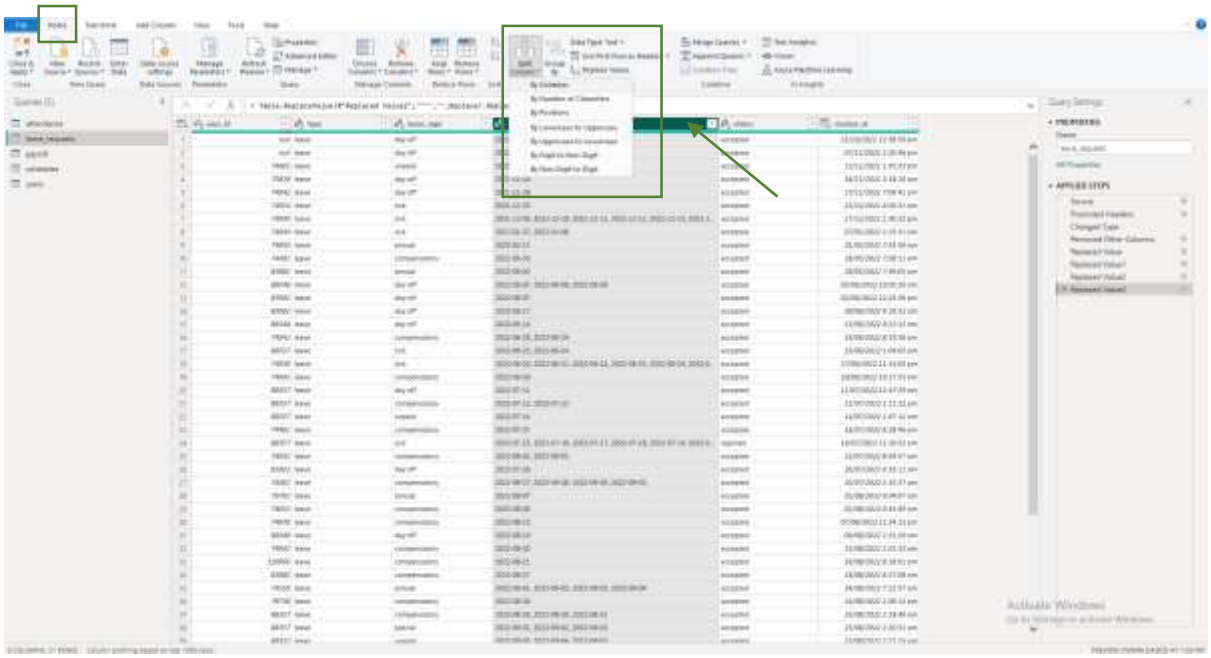
Right click again the cell of the dates column and select replace values. Input the double quote to Values to Find and leave the Replace With input box blank and then click ok.



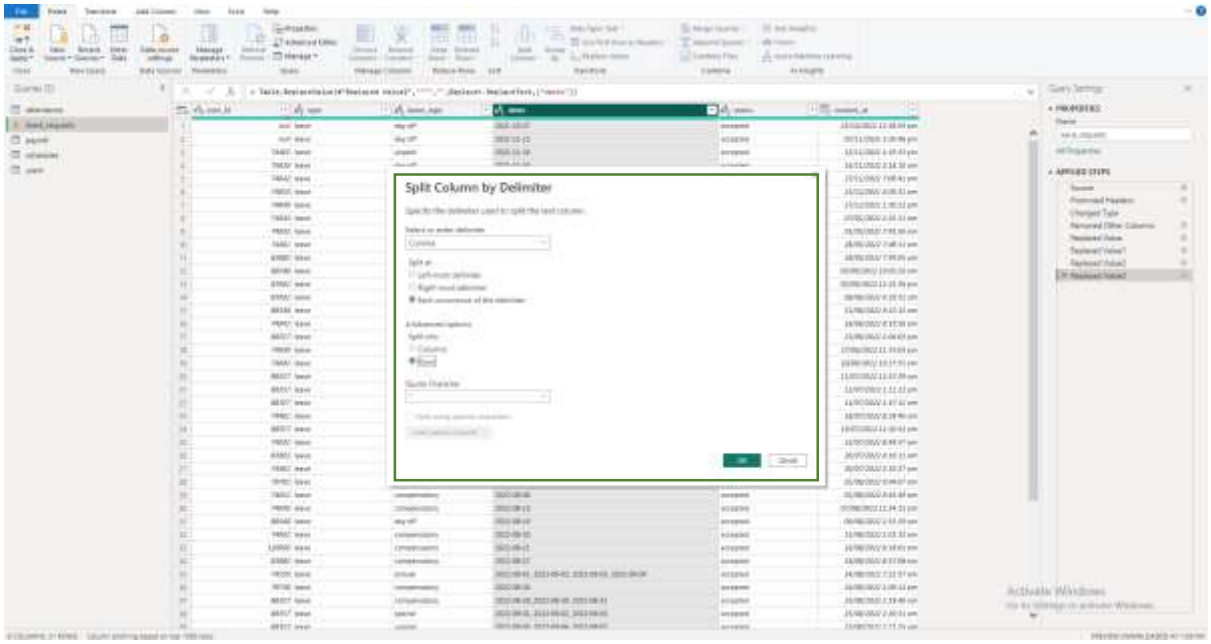
This is the result of removing the open and close brackets and double quotes.



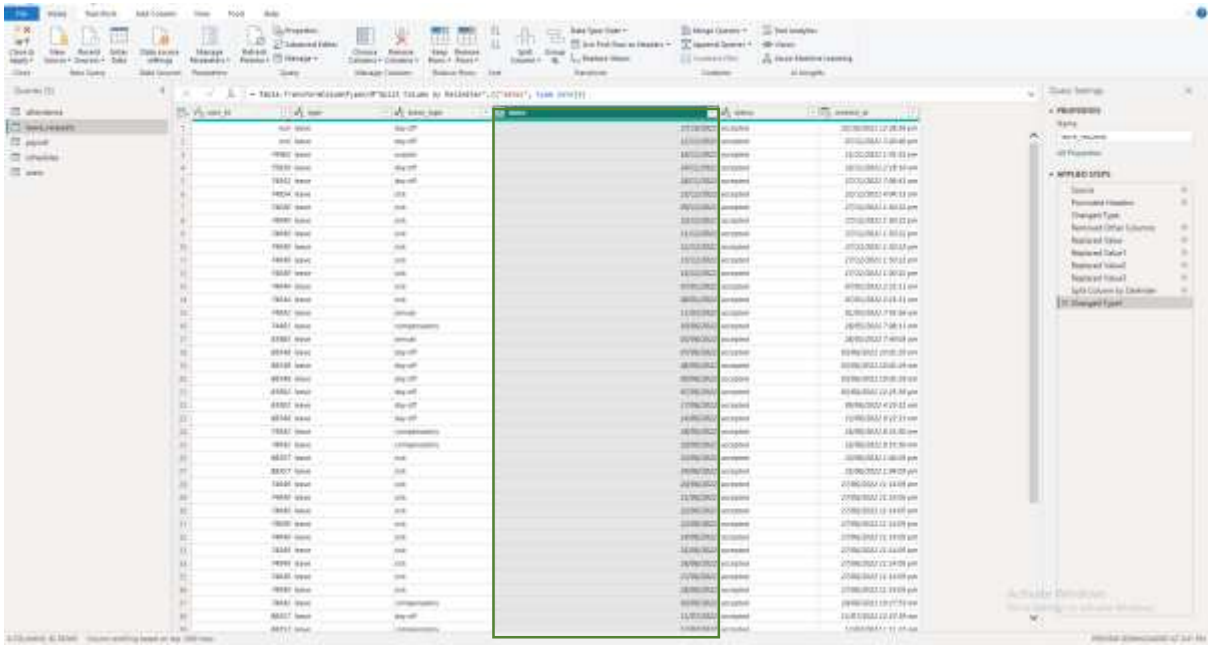
Splitting the date column by delimiter. Click to highlight the dates column, go to Home and click the Split Column dropdown arrow and select by Delimiter.



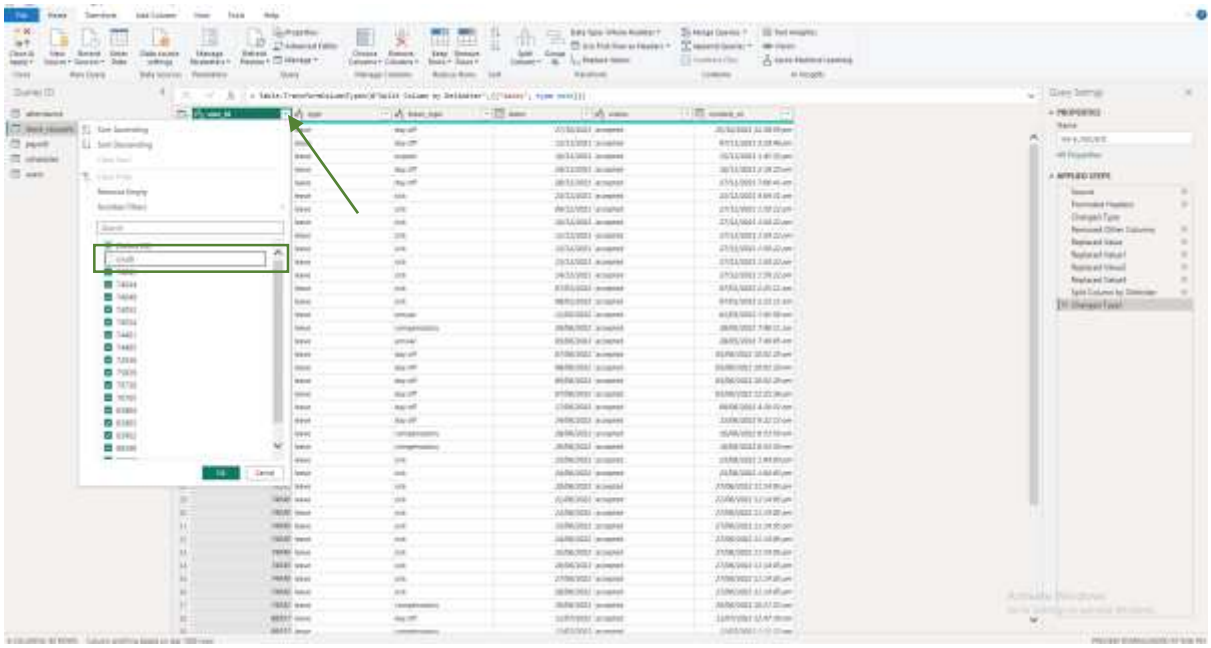
In the Split Column by delimiter pop up window, drop and select the “Comma” in the Select or enter delimiter. In the Split at choose “Each occurrence of the delimiter”. Click the Advance Options and in the split into choose “Rows" and then click ok.



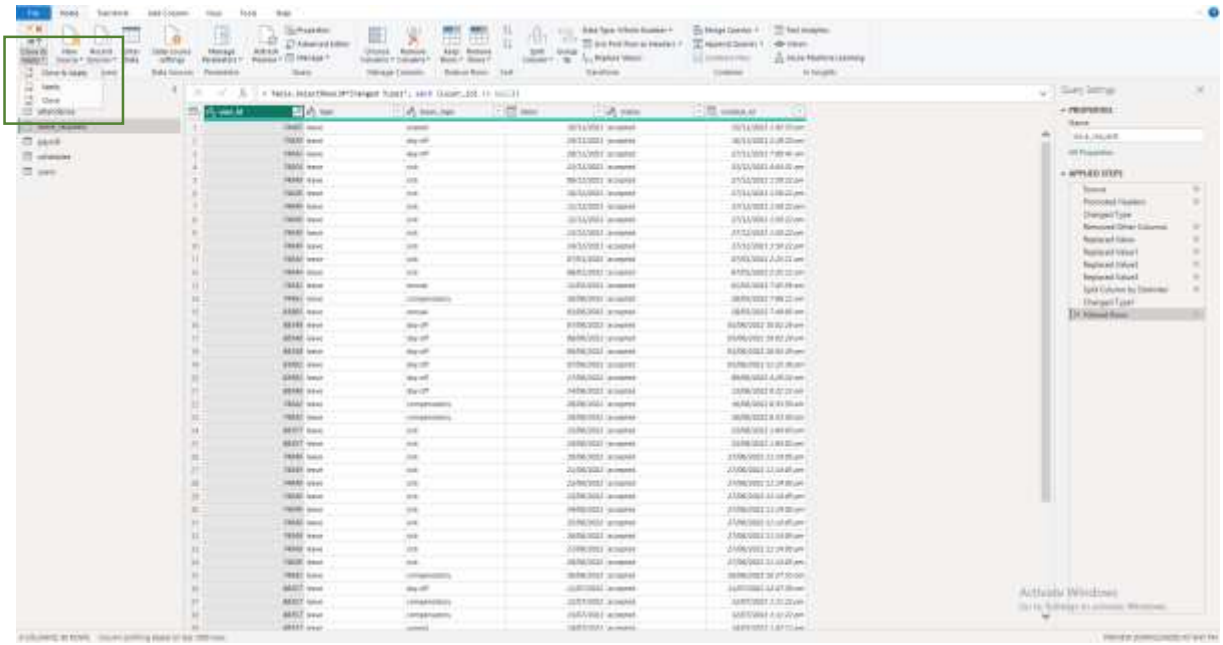
The result of Splitting the values of dates column by delimiter.



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.

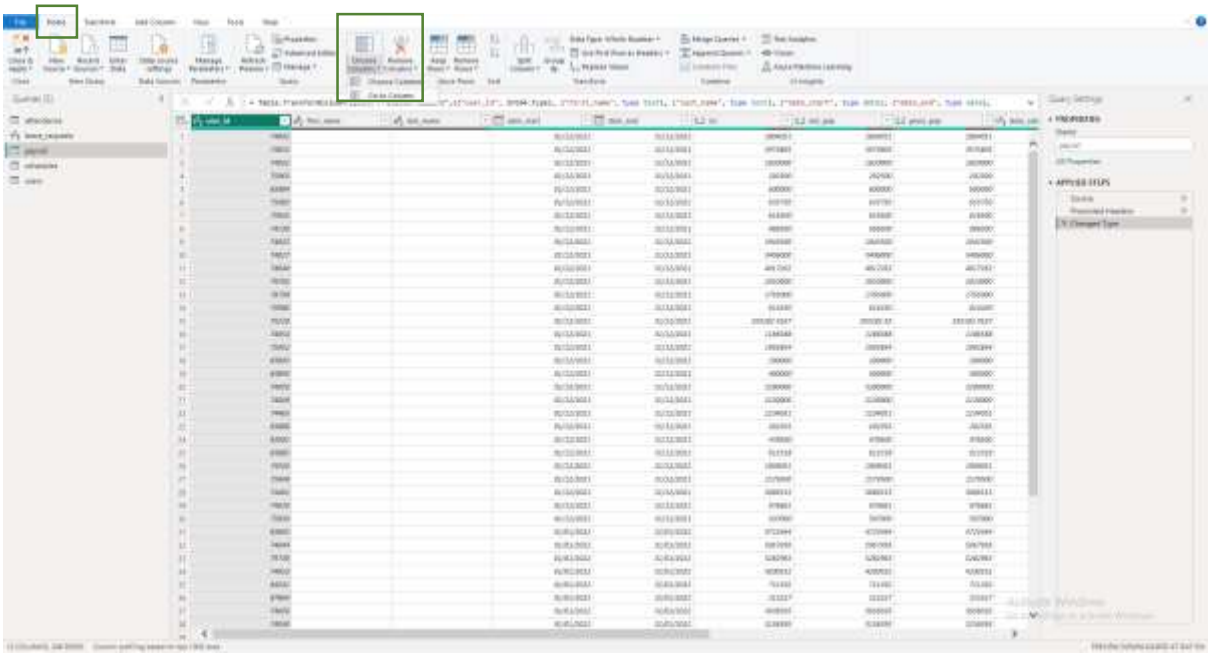


Go to and click Close and apply to load the data.The row count of leave_requests table is 90.

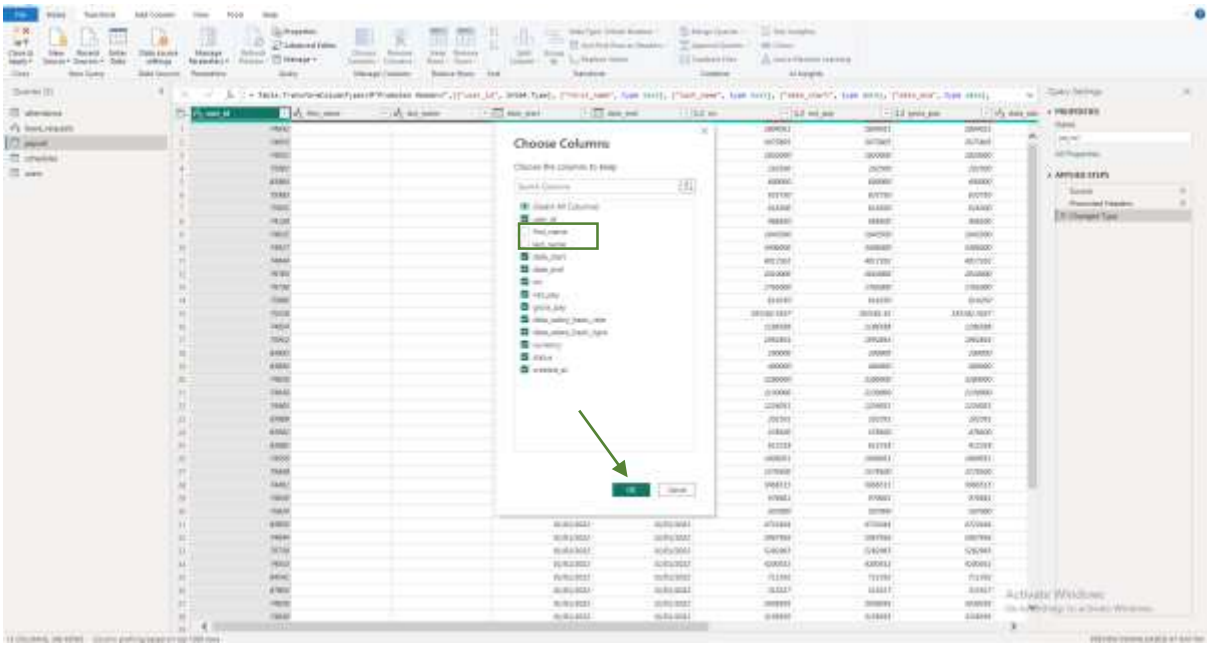


CLEANING AND TRANSFORMING PAYROLL TABLE

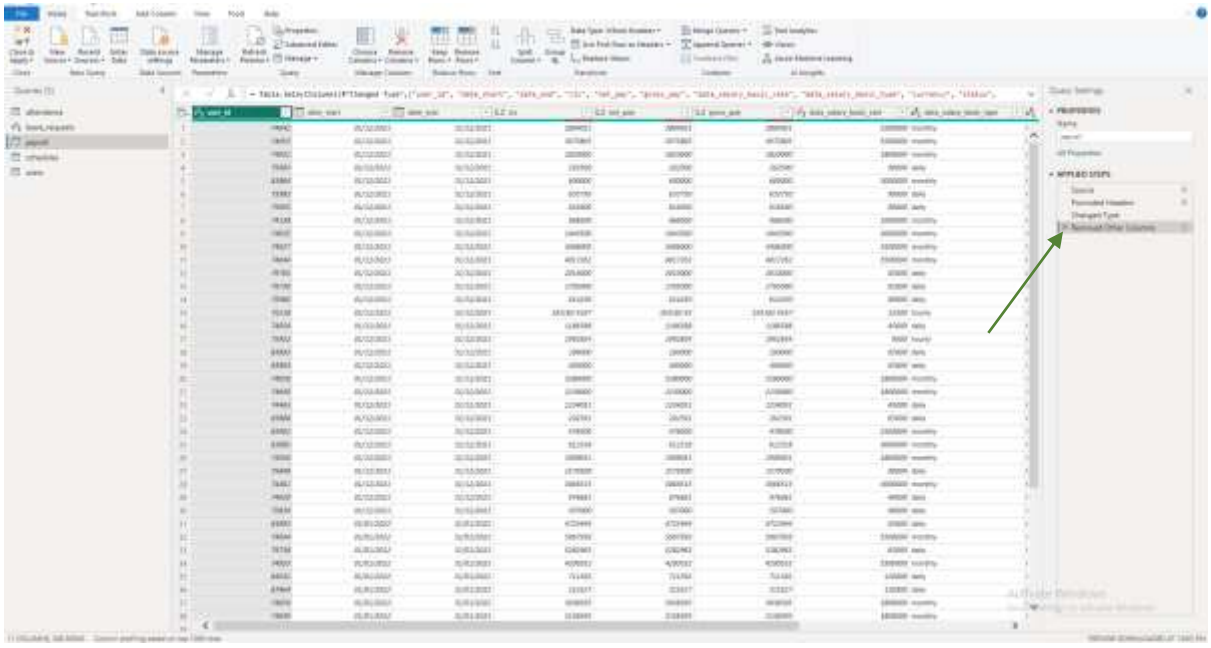
Remove unwanted columns from the payroll table. Go to Home and click the drop down arrow of Choose Columns and then select Choose Columns.



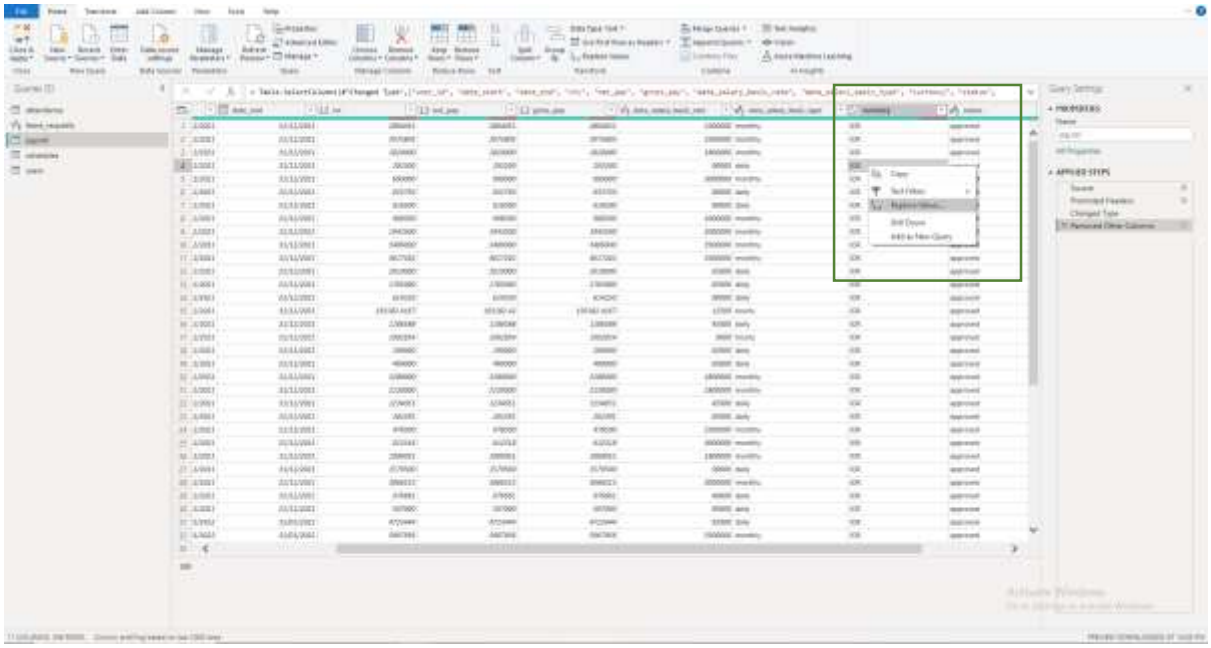
Unselect the check box of the column of first_name and last_name and then click ok.



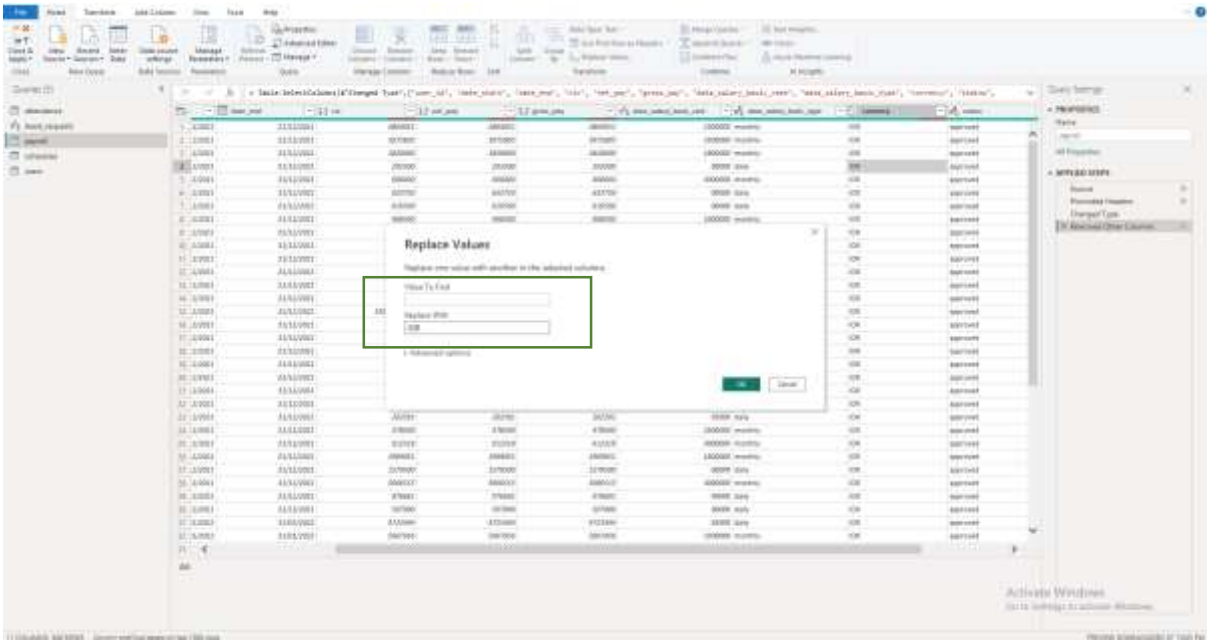
Result of removing columns first_name and last_name of payroll table.



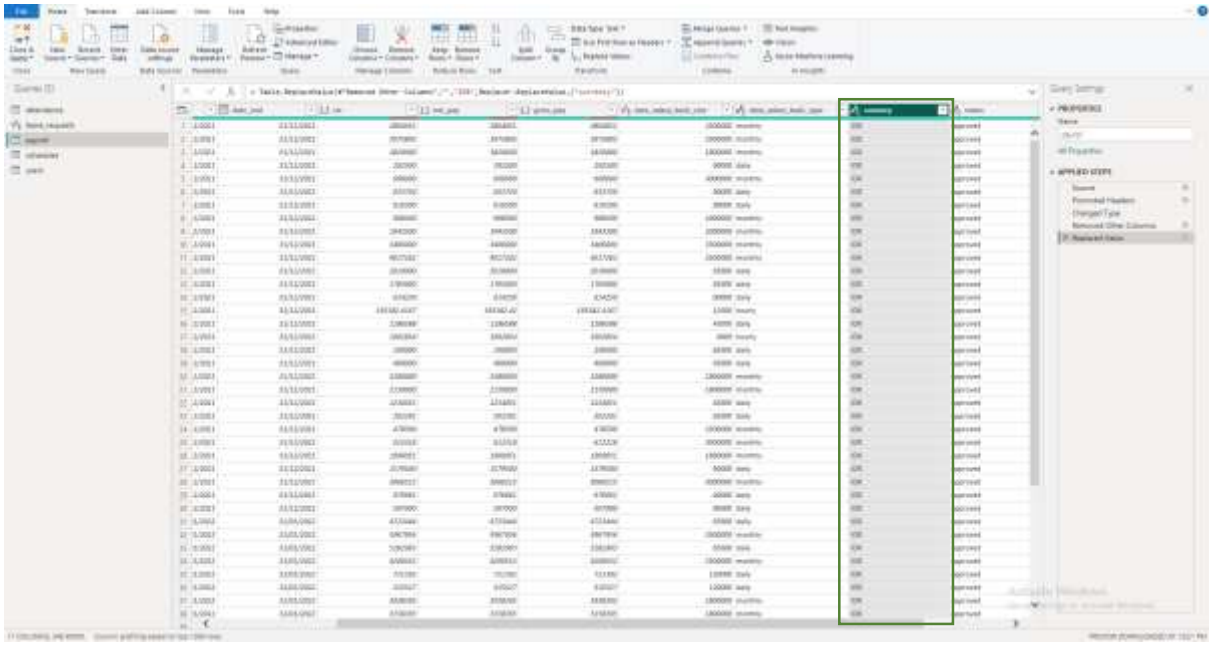
Populating the blank value of currency column with the IDR. Right Click the cell of the currency column and select the replace values.



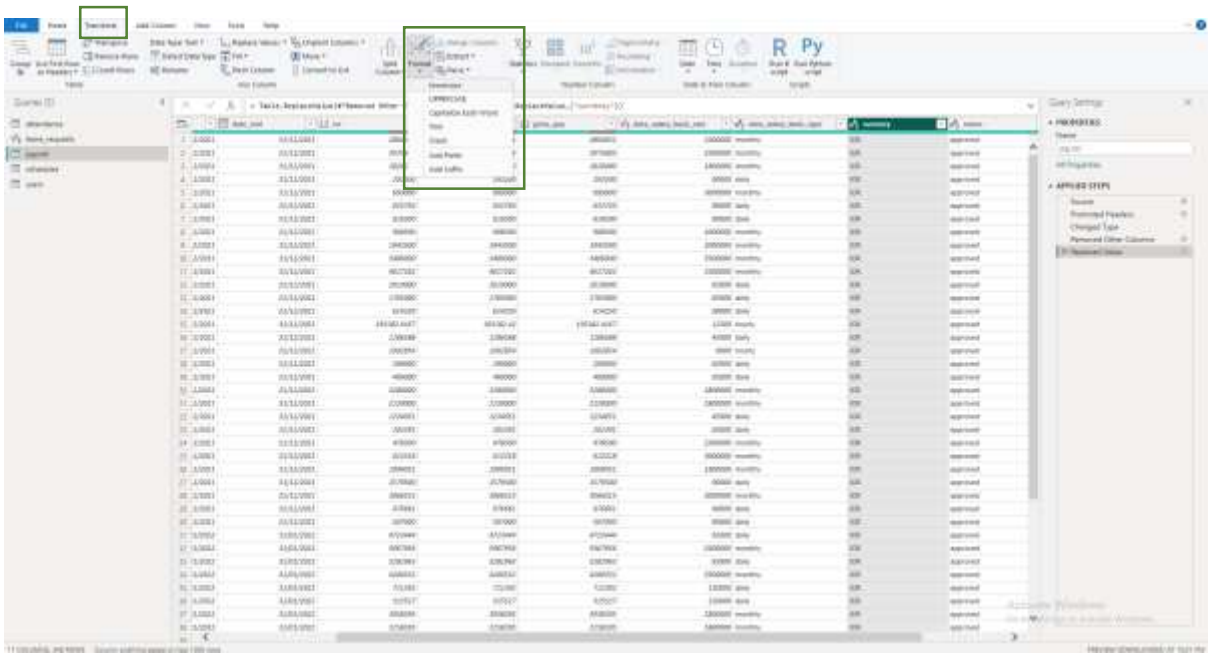
Leave the Value to Find blank, put IDR to the input box of Replace With and then click ok.



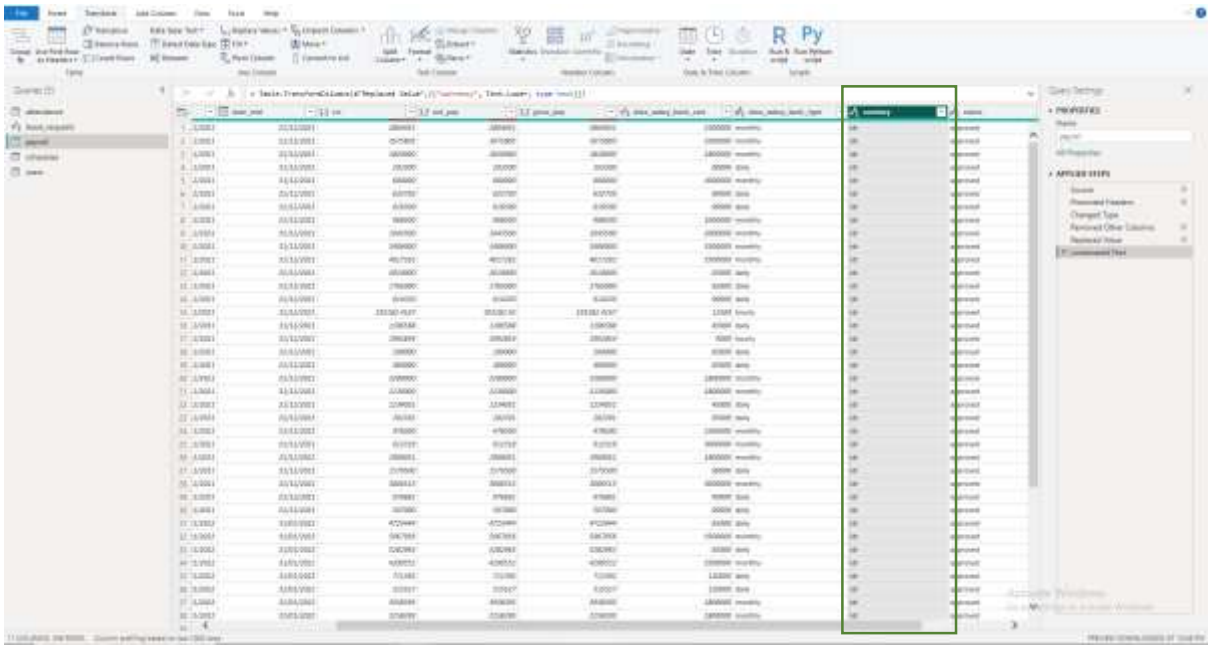
Result of populating the blank rows of currency column.



Changing the text format of currency column to lowercase letter. Highlight the currency column.Go to Transform, select and click the drop down arrow of Format and click lowercase.

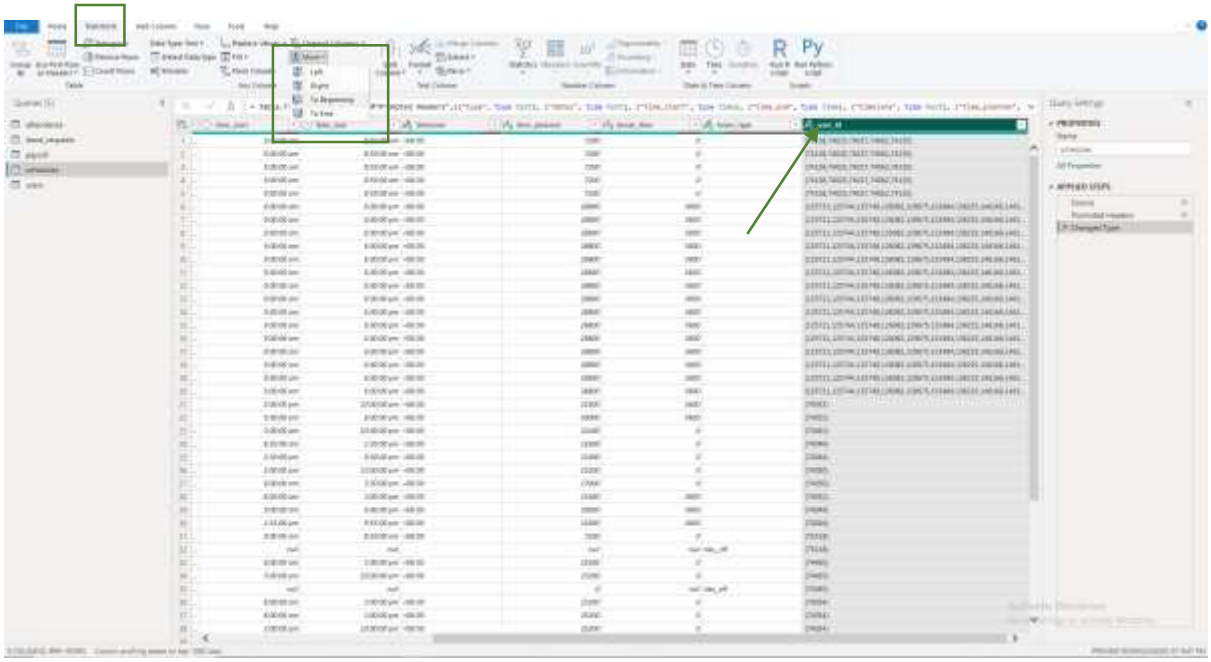


Result of changing text format of currency column to lowercase. The count of rows in the payroll table is 348.

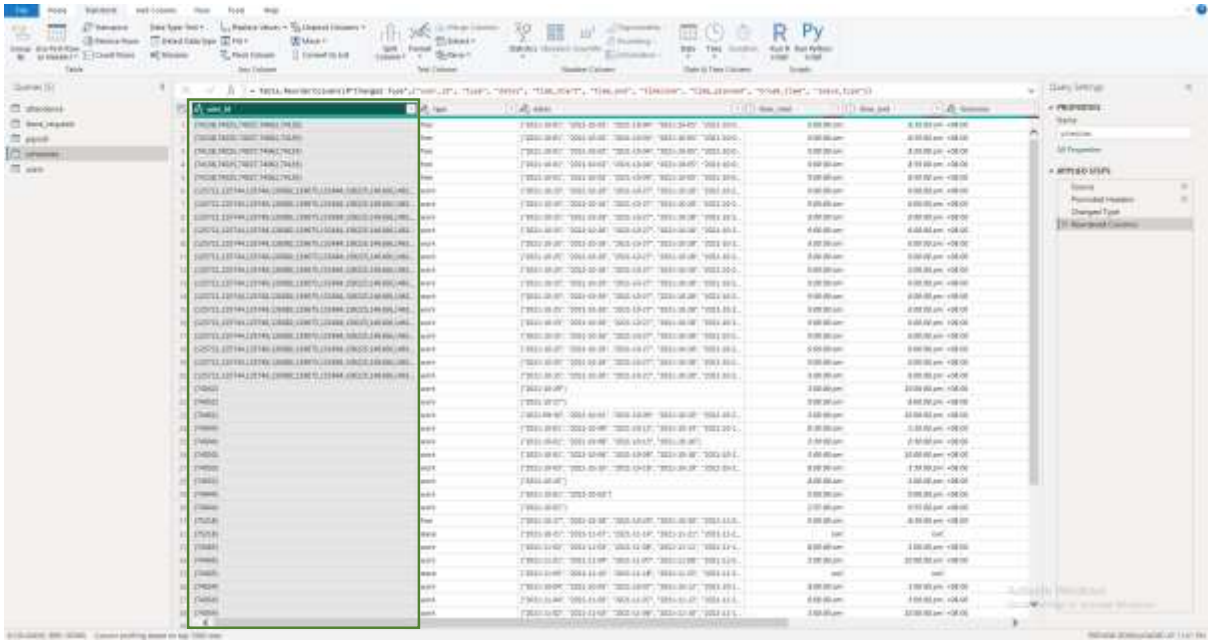


CLEANING AND TRANSFORMING SCHEDULES TABLE

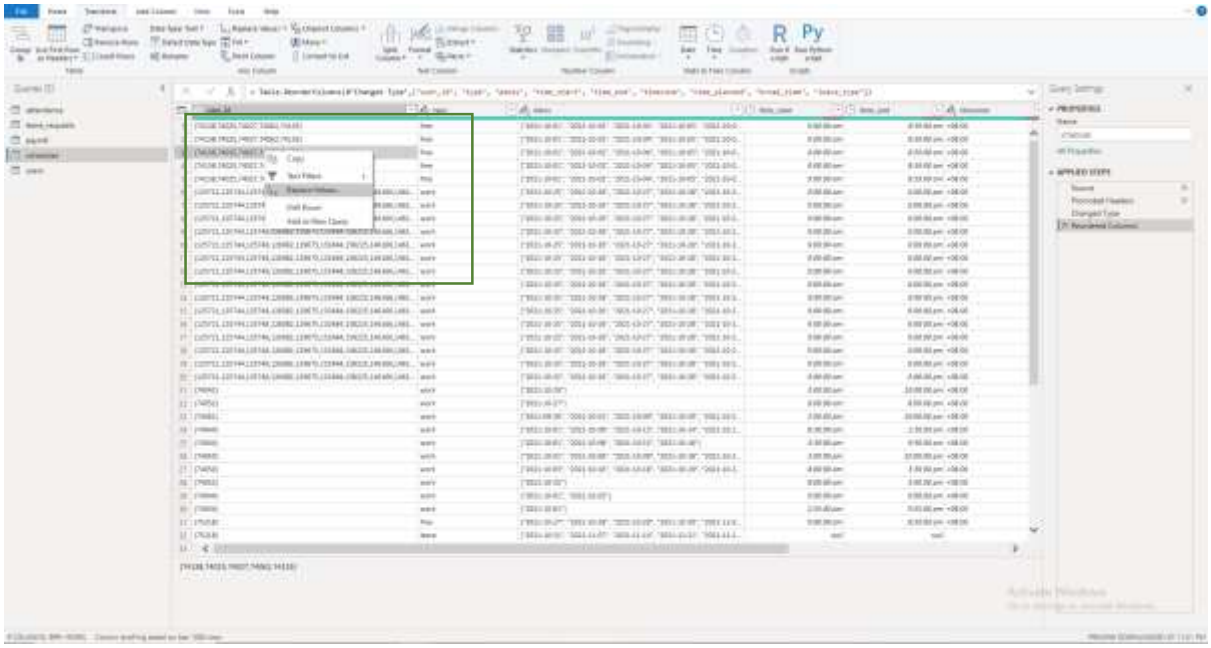
Transferring the user_id column to the beginning of the table. Click to highlight the user_id column. Go to Transform, click the drop down arrow of Move then select To Beginning.



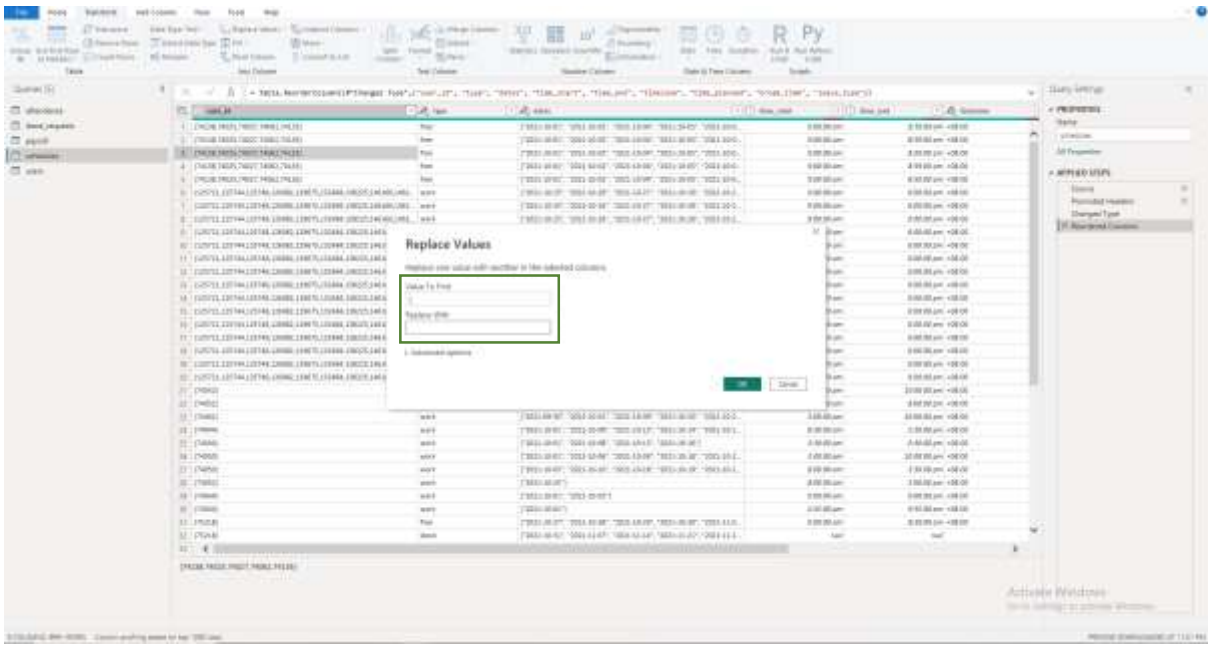
Result of schedules table column transfers of user_id.



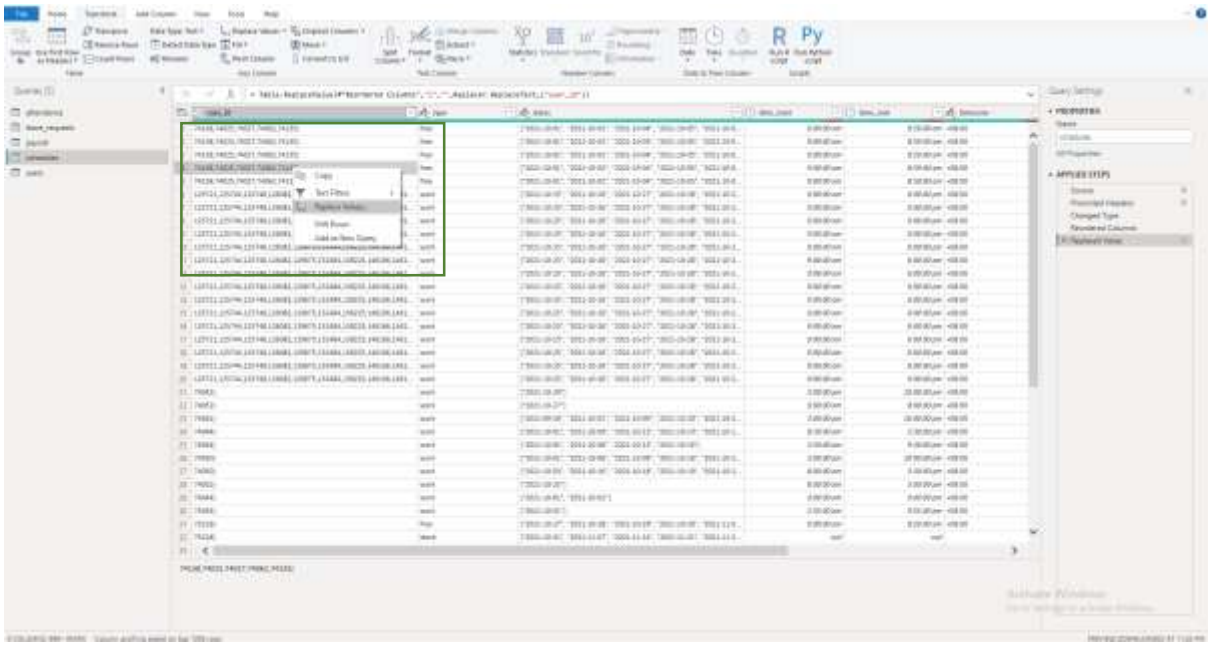
Cleaning the user_id column of schedules table. Removing the curly brackets, right click the cell of the user_id column and click replace values.



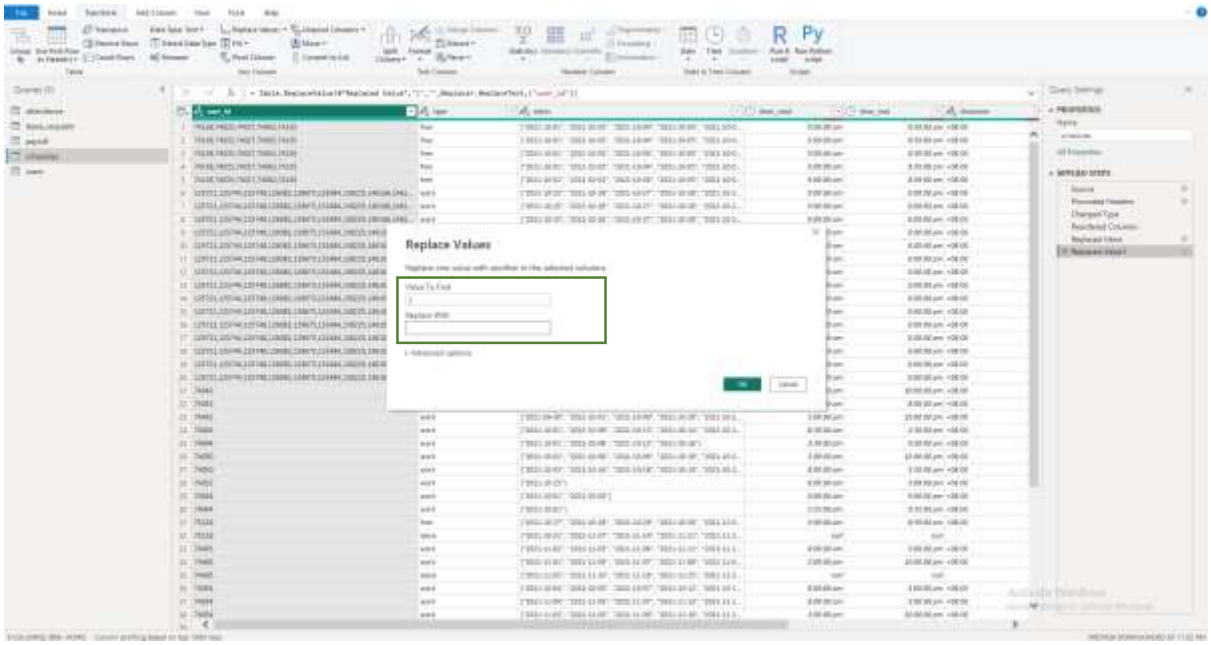
Input an open curly bracket to Values to Find, leave the Replace With input box blank and then click ok.



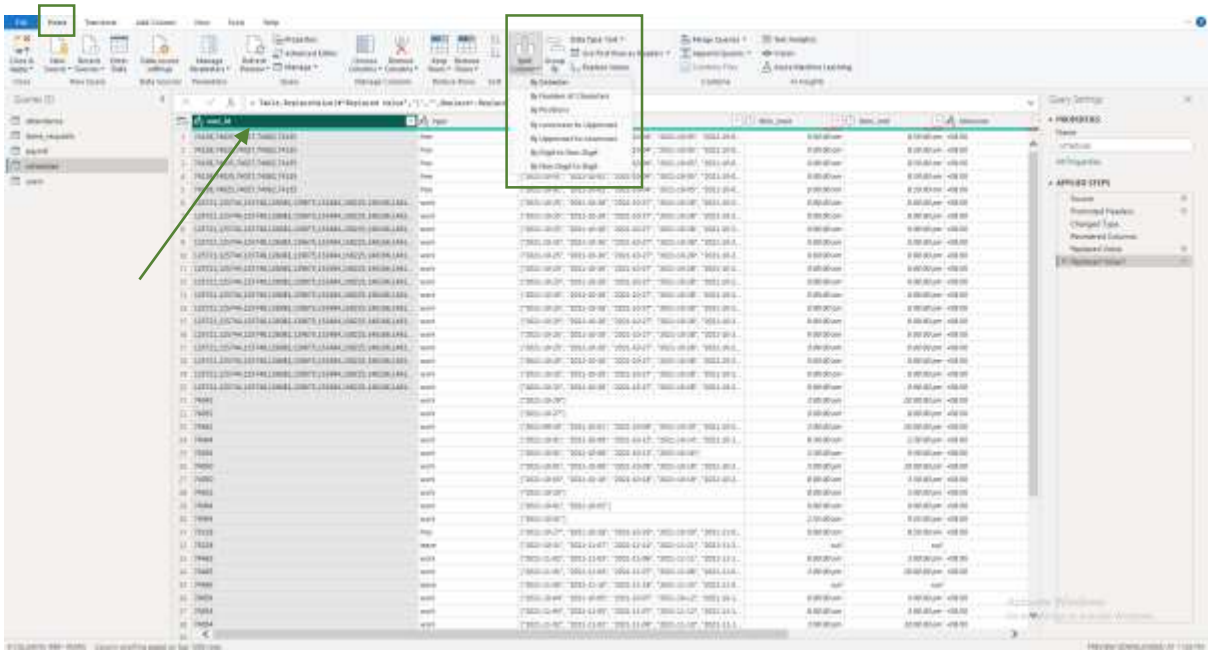
Right click again the user_id column cell and click the replace values.



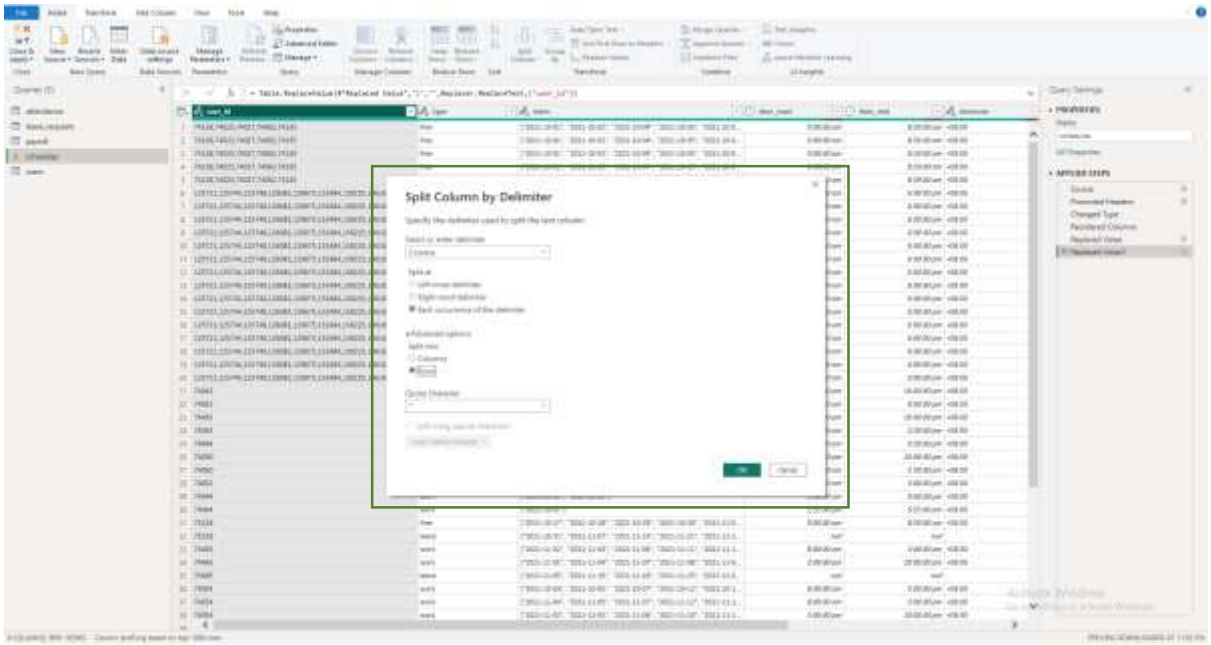
Input a close curly bracket to Values to Find, leave the Replace With input box blank and then click ok.



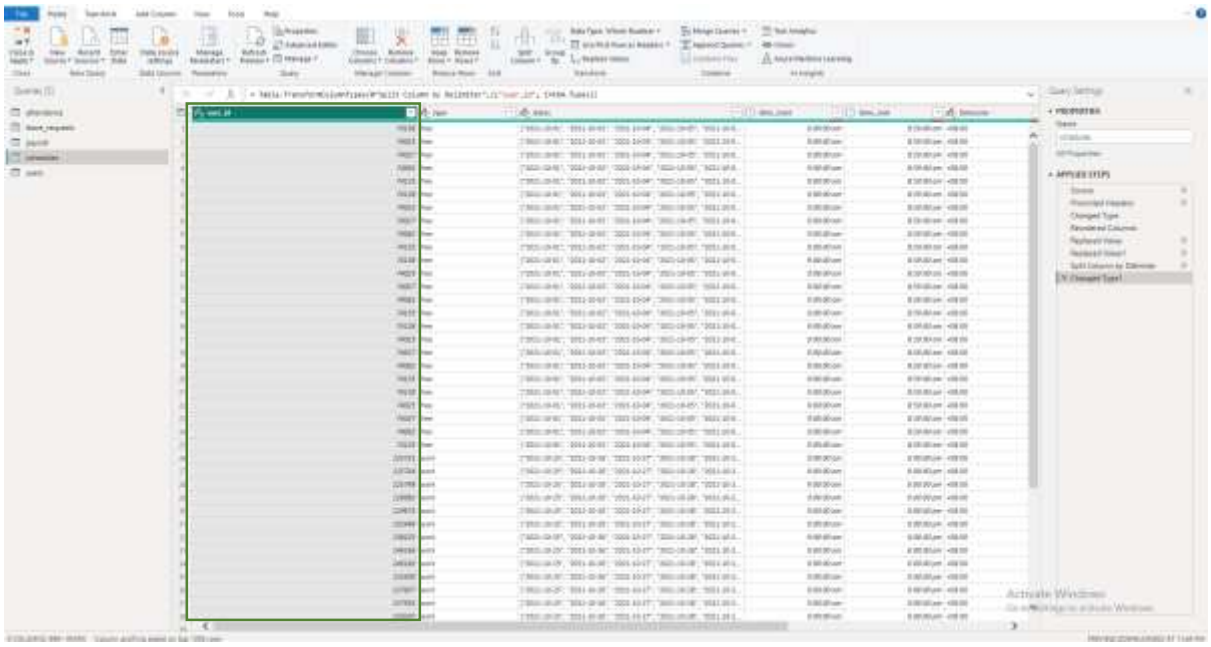
Splitting the value by delimiter of the user_id. Click to highlight the user_id column, go to Home and click the Split Column drop down arrow then select by Delimiter.



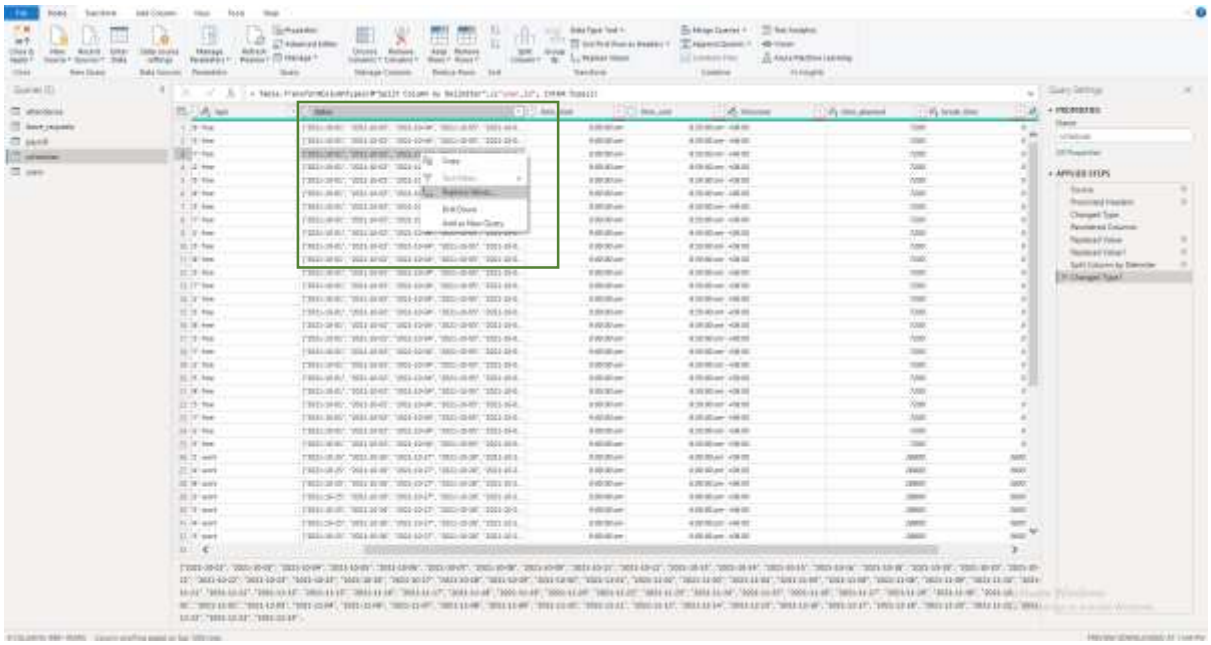
In the Split Column by Delimiter window, click the drop down arrow of Select or Enter delimiter and choose Comma. In the Split at choose Each occurrence of the delimiter, click Advance options and choose Rows and then click ok.



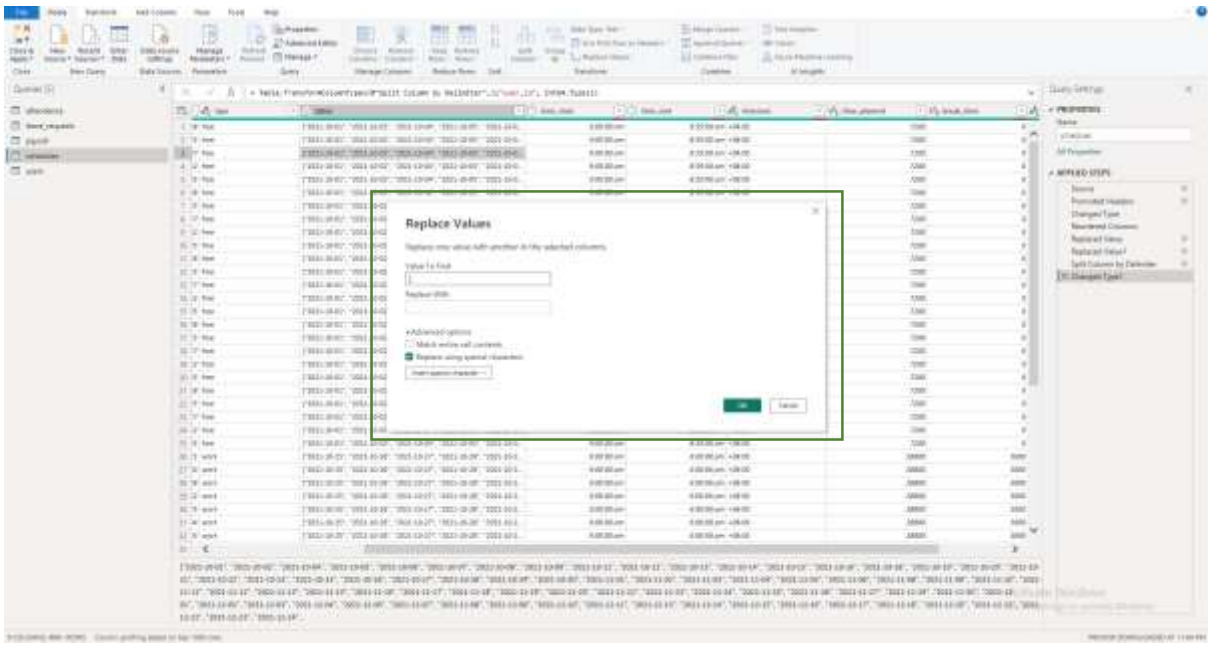
Result of splitting the values of user_id column by delimiter.



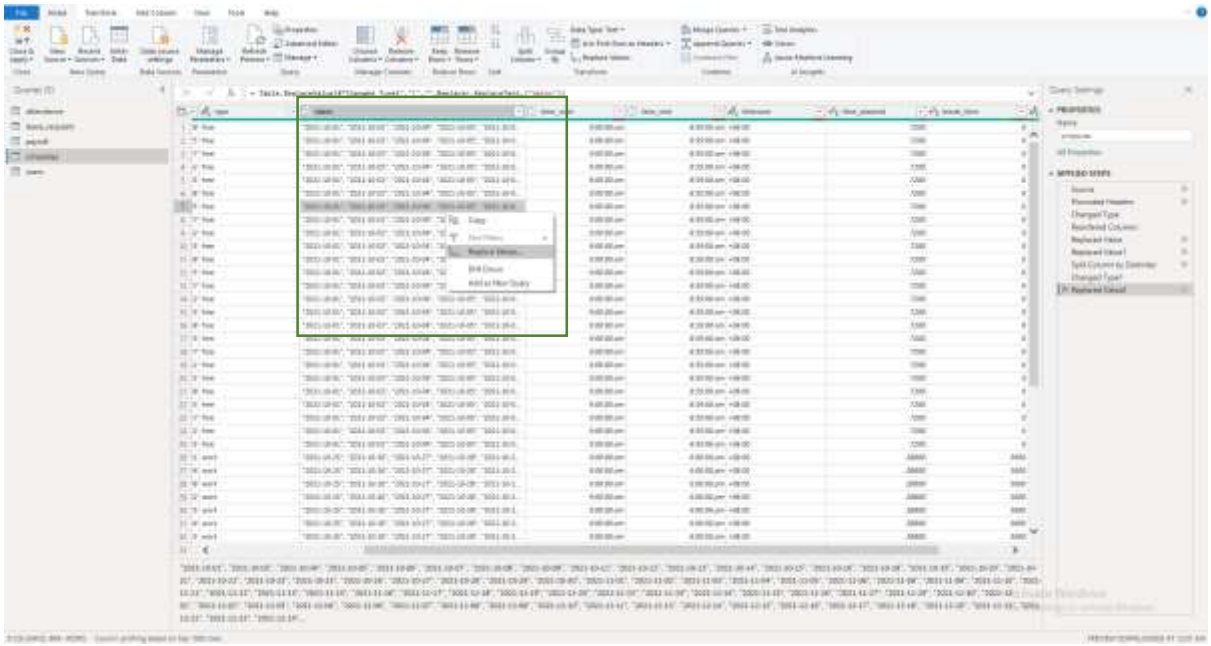
Cleaning the dates column of schedules table. Right click the cell of dates column and select replace values.



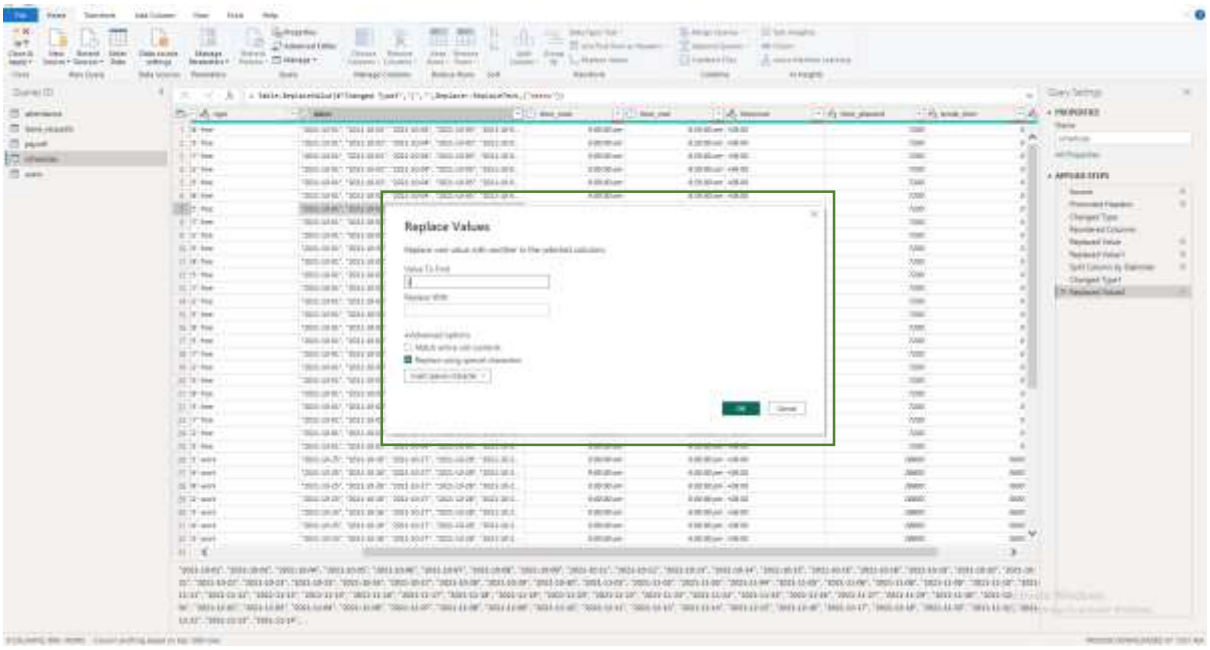
Input an open bracket to Values to Find, leave the Replace With input box blank and then click ok.



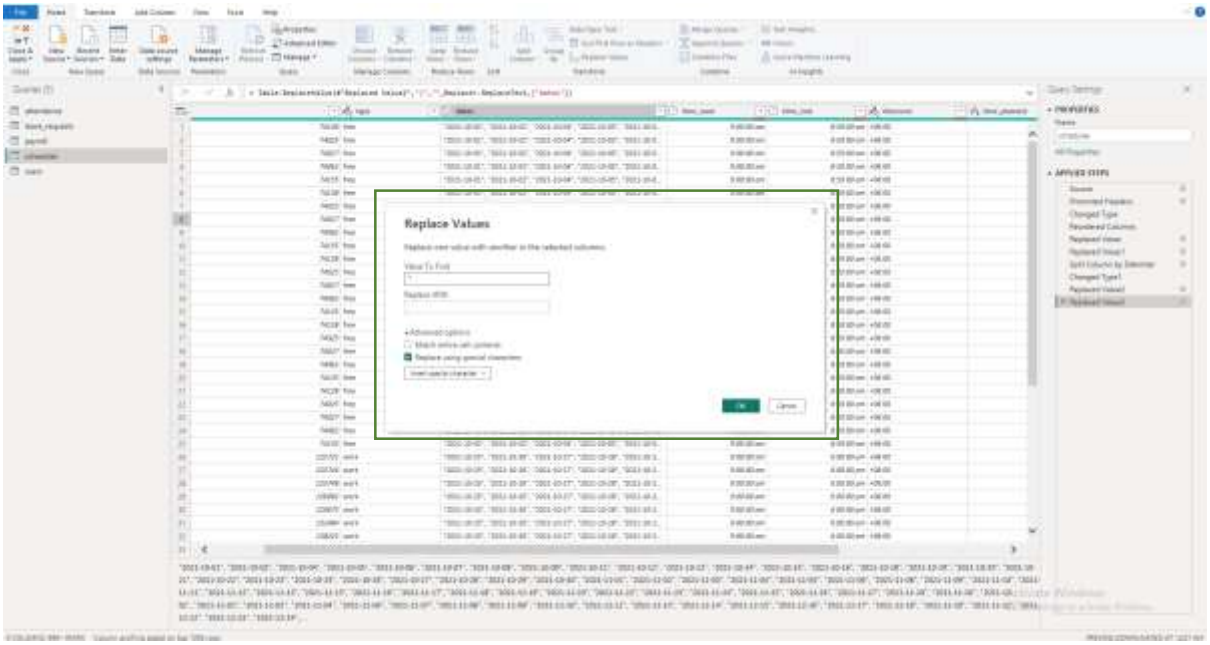
Right click again the cell of dates column and select replace values.



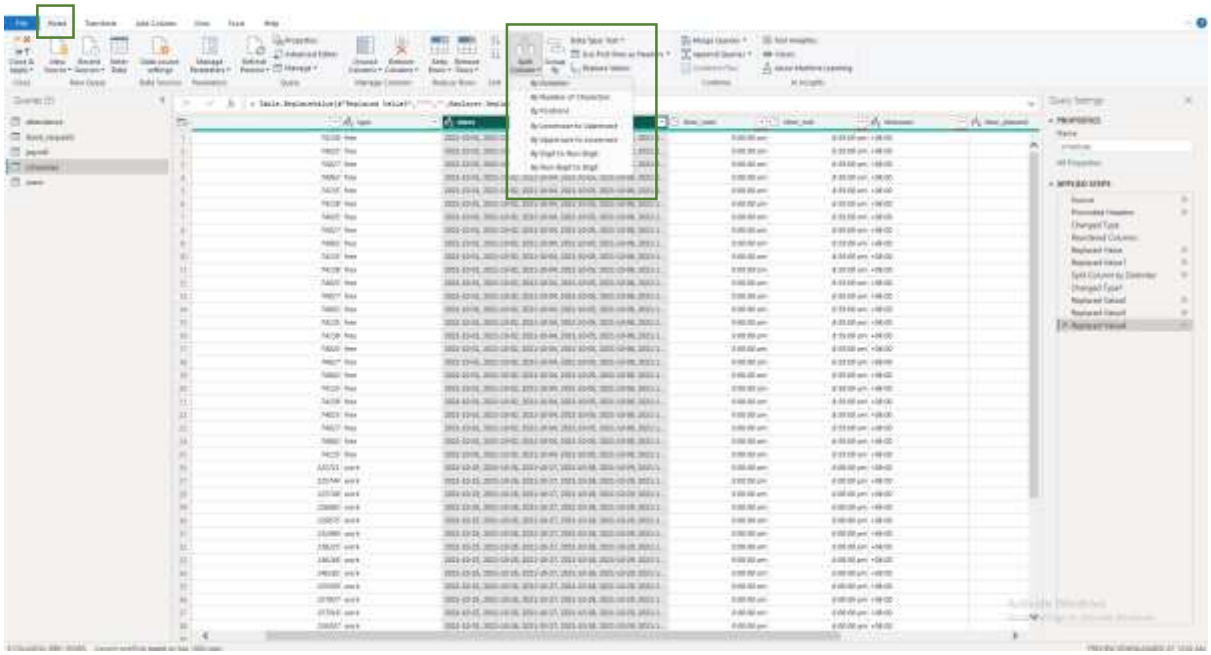
Input a close bracket to Values to Find, leave the Replace With input box blank and then click ok.



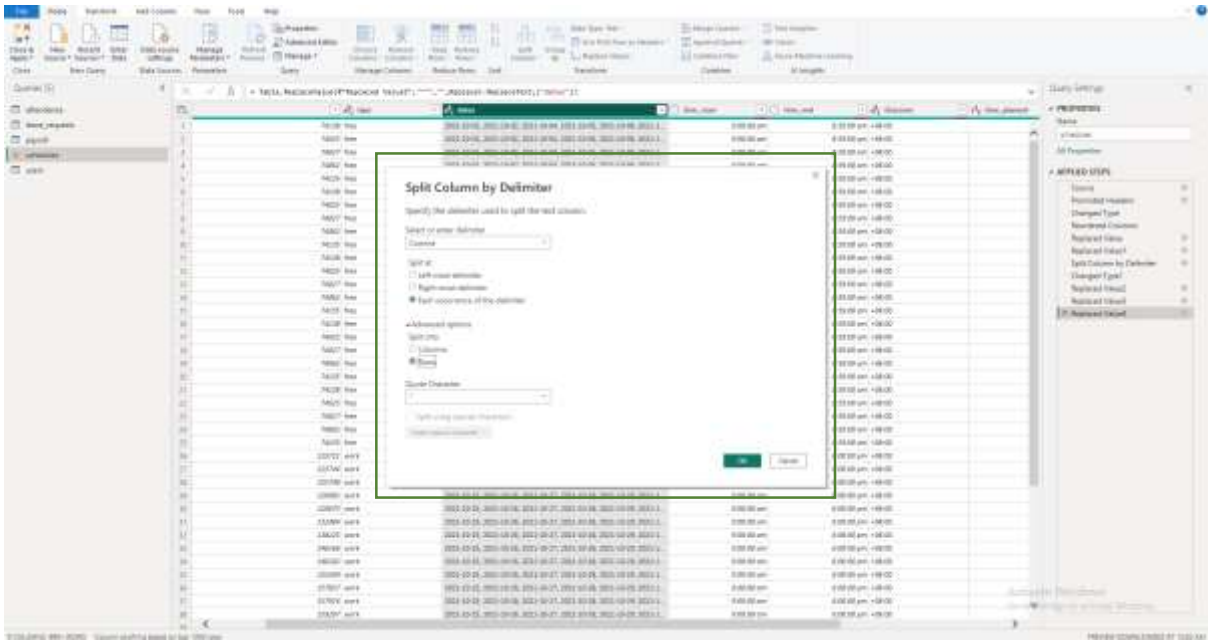
Right click again the cell of dates column and select replace values. Input an open double quote to Values to Find, leave the Replace With input box blank and then click ok.



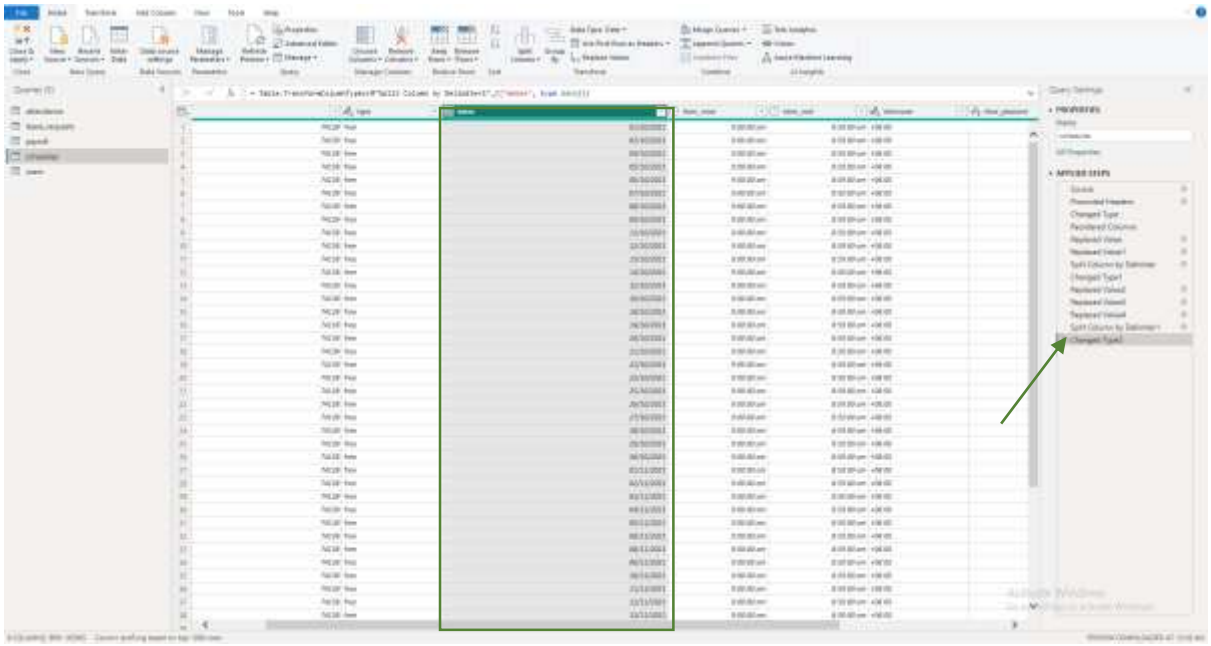
Splitting the dates column by delimiter. Click to highlight the dates column, go to Home and click the Split Column dropdown arrow and select by Delimiter.



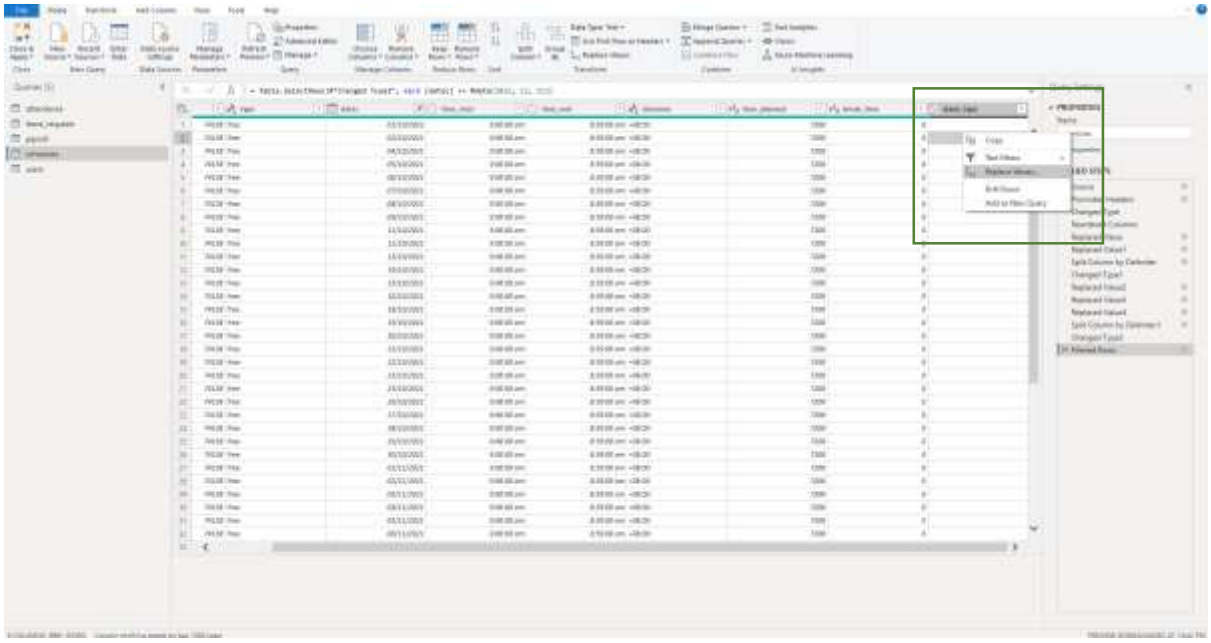
In the Split Column by Delimiter window, click the drop down arrow of Select or Enter delimiter and choose Comma. In the Split at choose Each occurrence of the delimiter, click Advance options and choose Rows and then click ok.



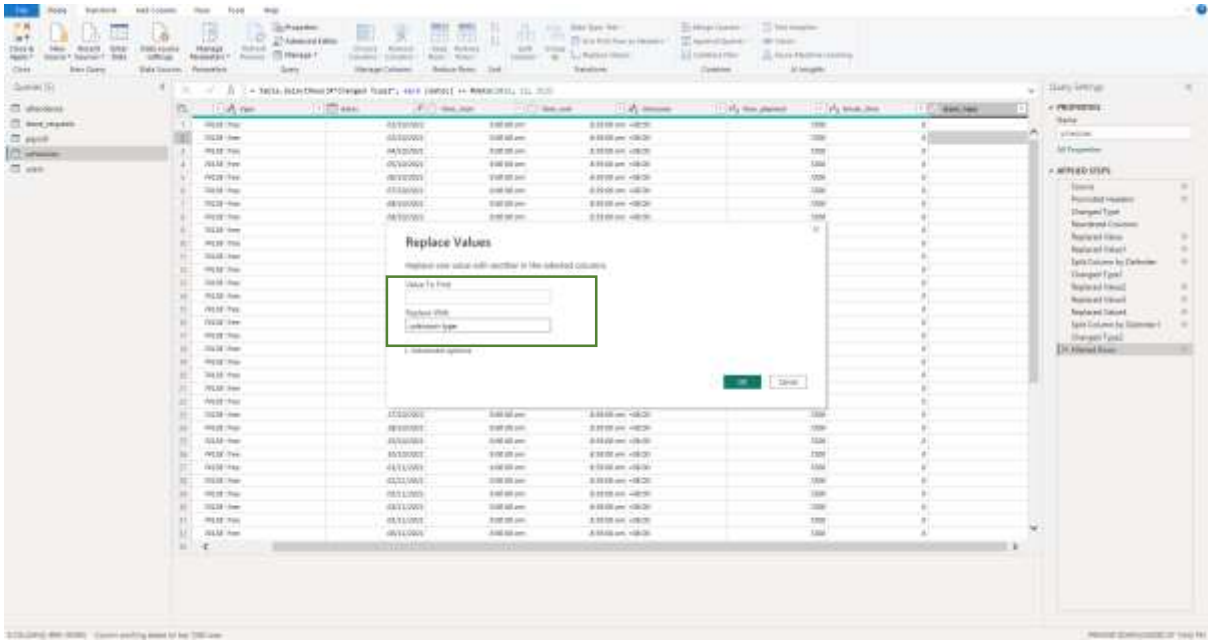
Result of schedules table splitting the dates column by delimiter.



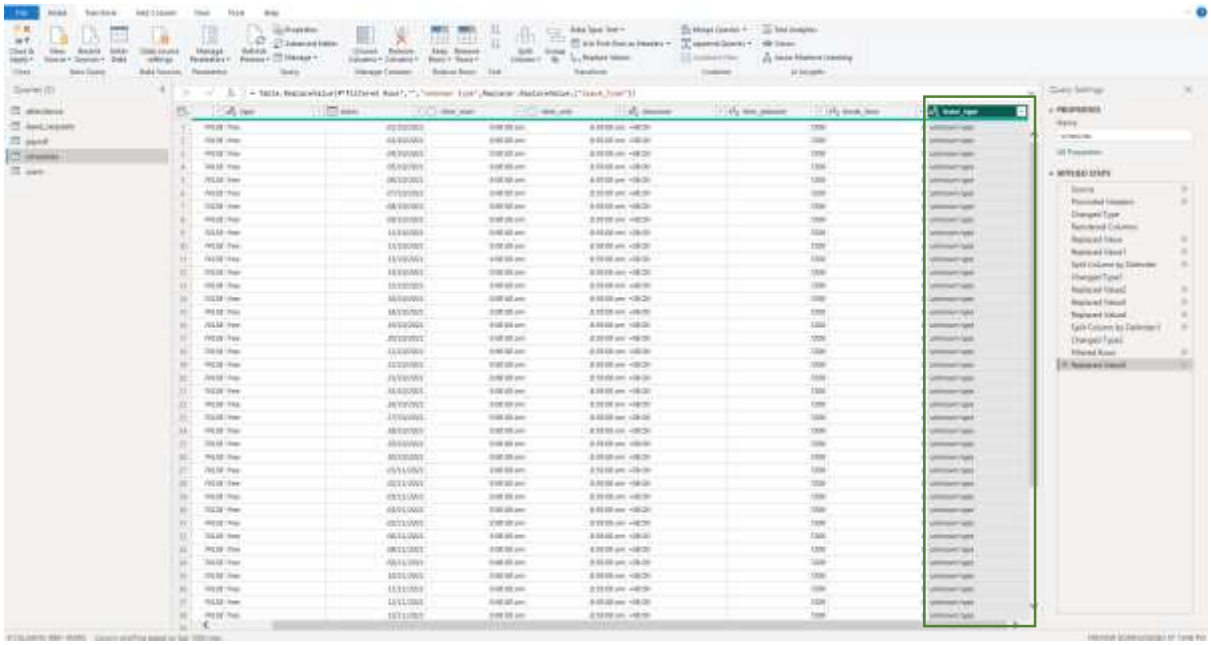
Replacing the blank value in the leave_type column with unknown type in the schedules table. Right click the cell of leave_type column and click replace values.



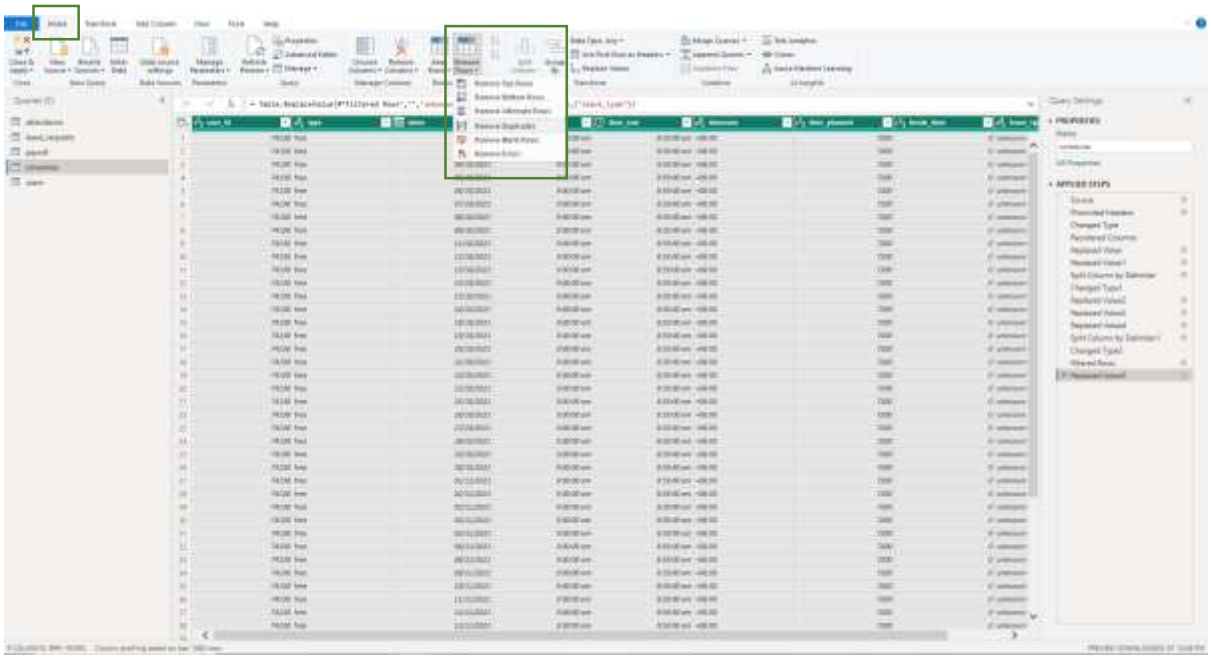
Leave the input box of Value to Find blank, input unknown type to the Replace With input box and then click ok.



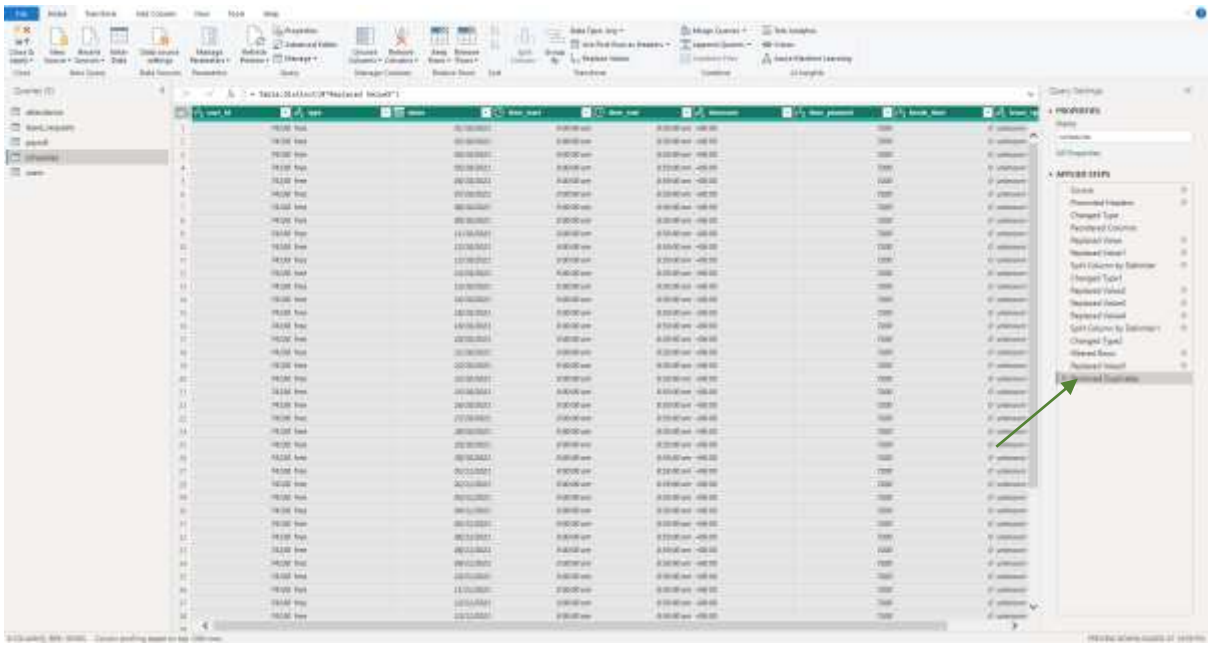
Result of replacing blank value in the leave_type column with unknown type



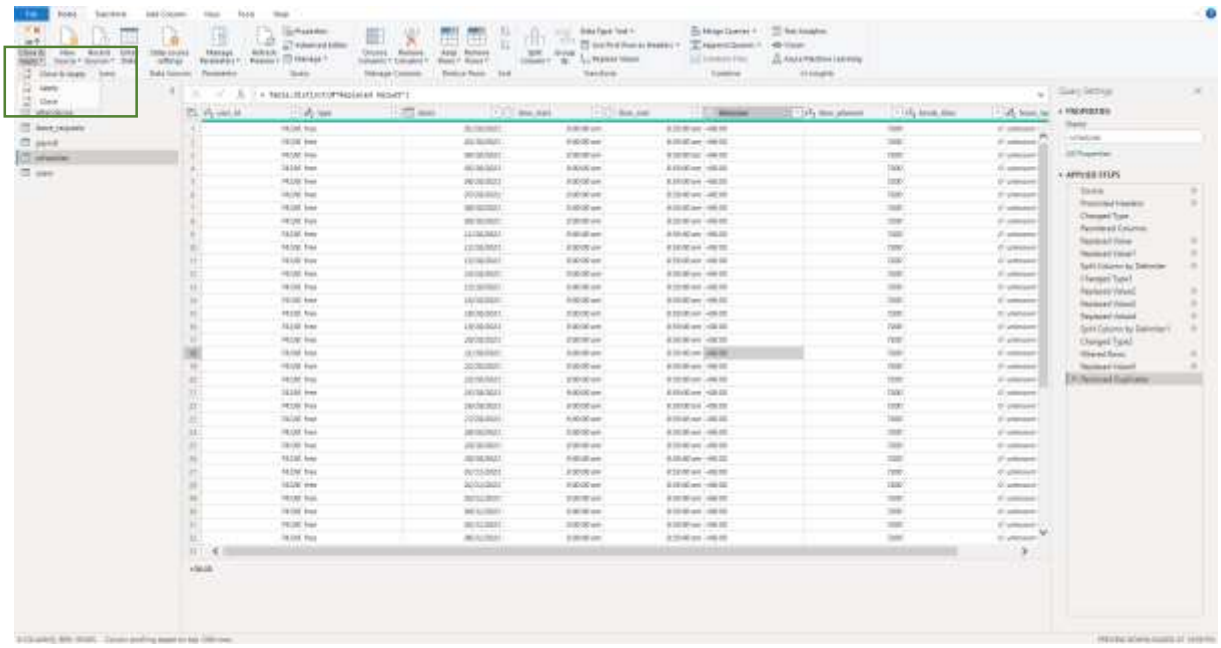
Removing duplicates values in schedules table. Highlight all the column of schedules table. Go to Home, select and click the drop down arrow of Remove Rows and then click Remove Duplicates.



Result of removing duplicate values of schedules table

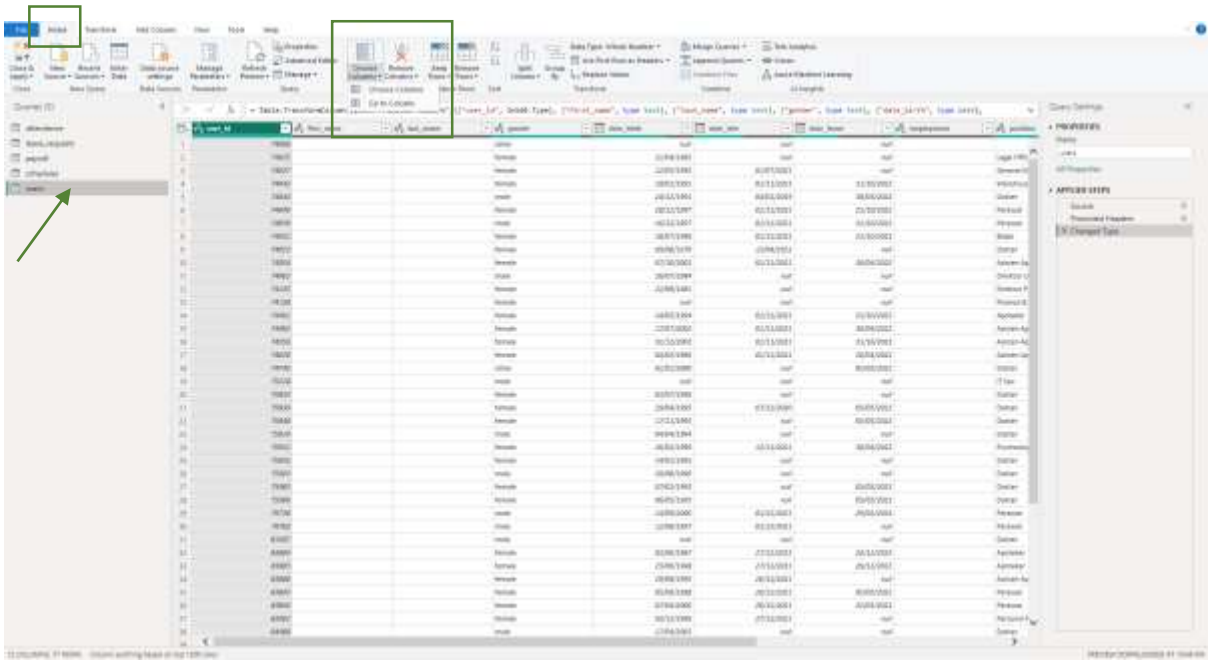


Total of 21676 rows remain in the schedule table after removing duplicate rows. Click close and apply to load the data.

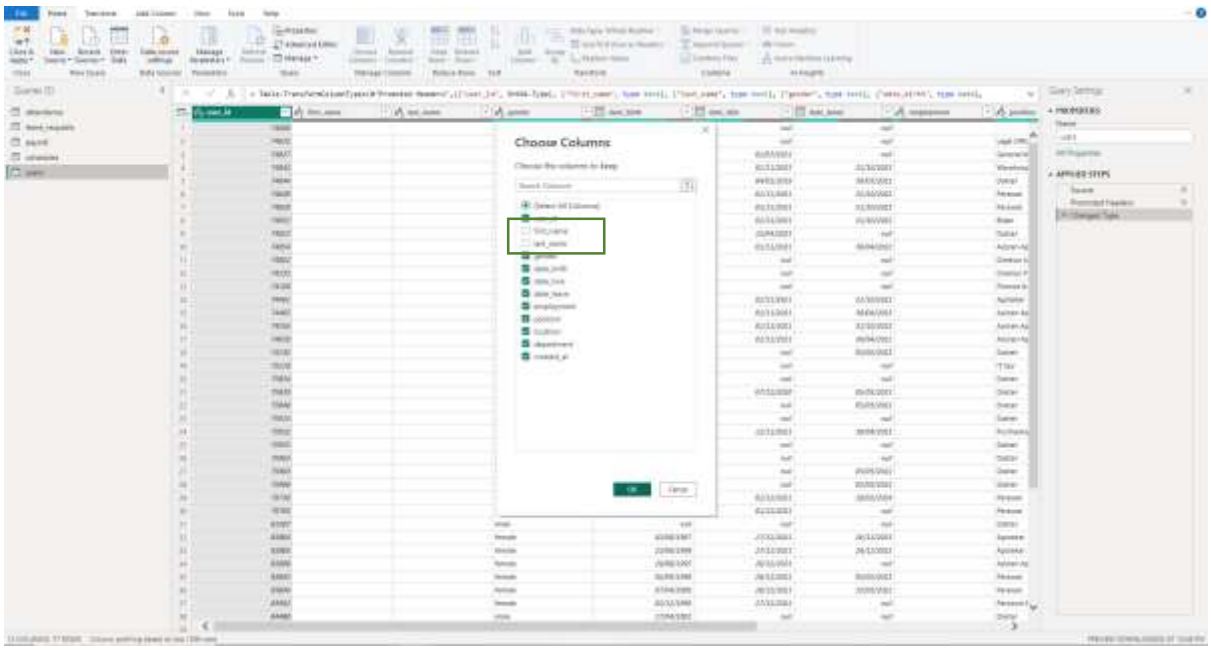


CLEANING AND TRANSFORMING USERS TABLE

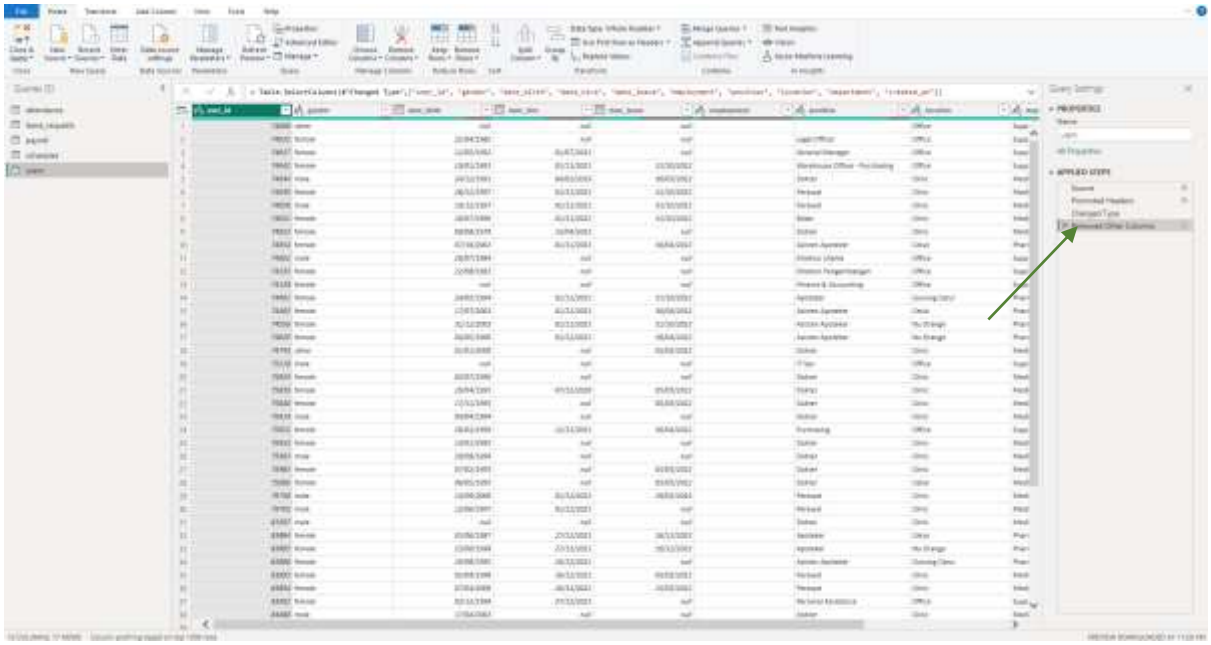
Removing unwanted column in users table. Select the users table in the Navigator Pane. Go to Home, click the drop down arrow of Choose Columns and click Choose Columns.



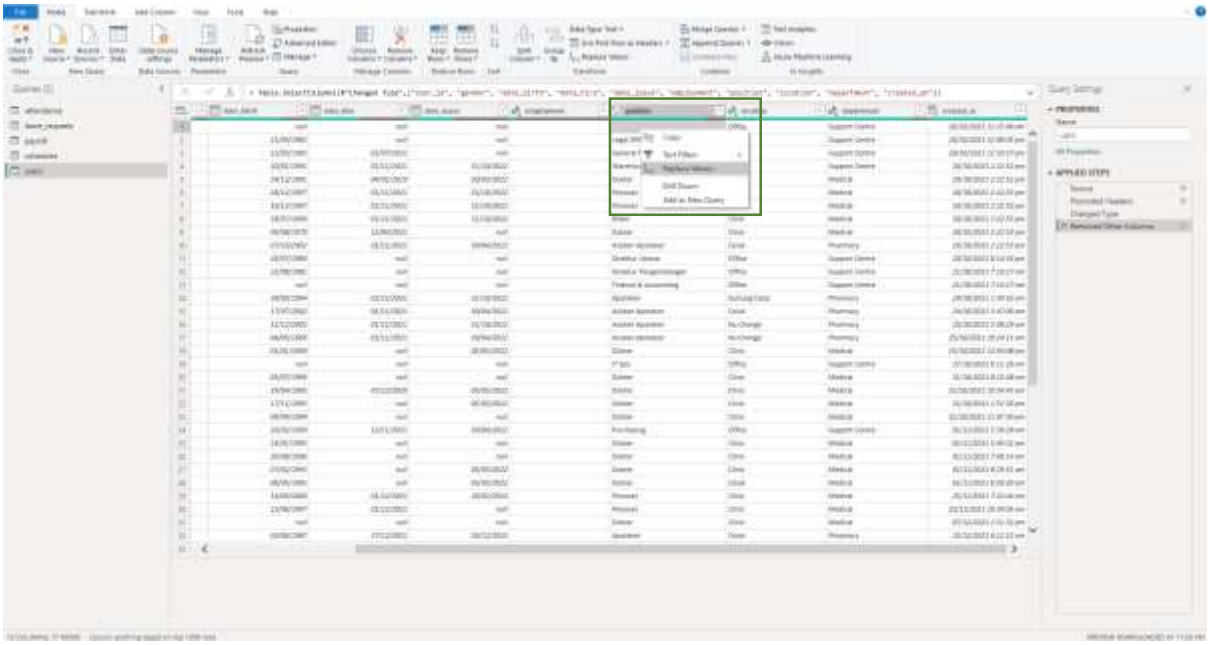
Uncheck the box of column you want to remove from the table, we choose first_name and last_name and then click ok.



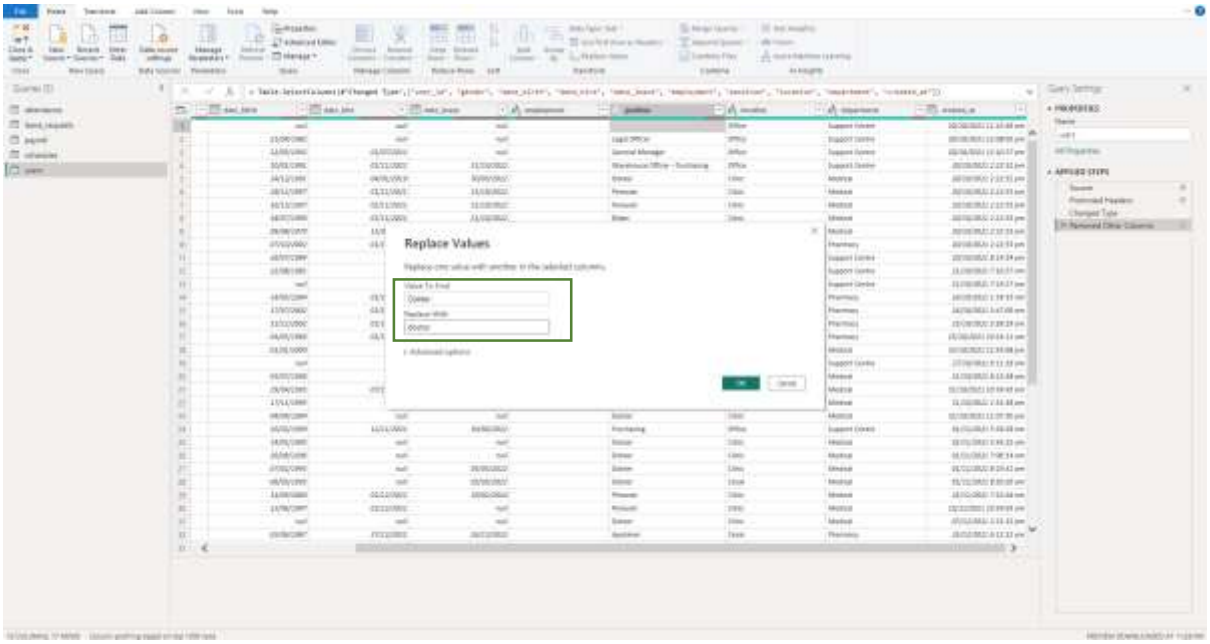
Result of removing unwanted columns from the users table.



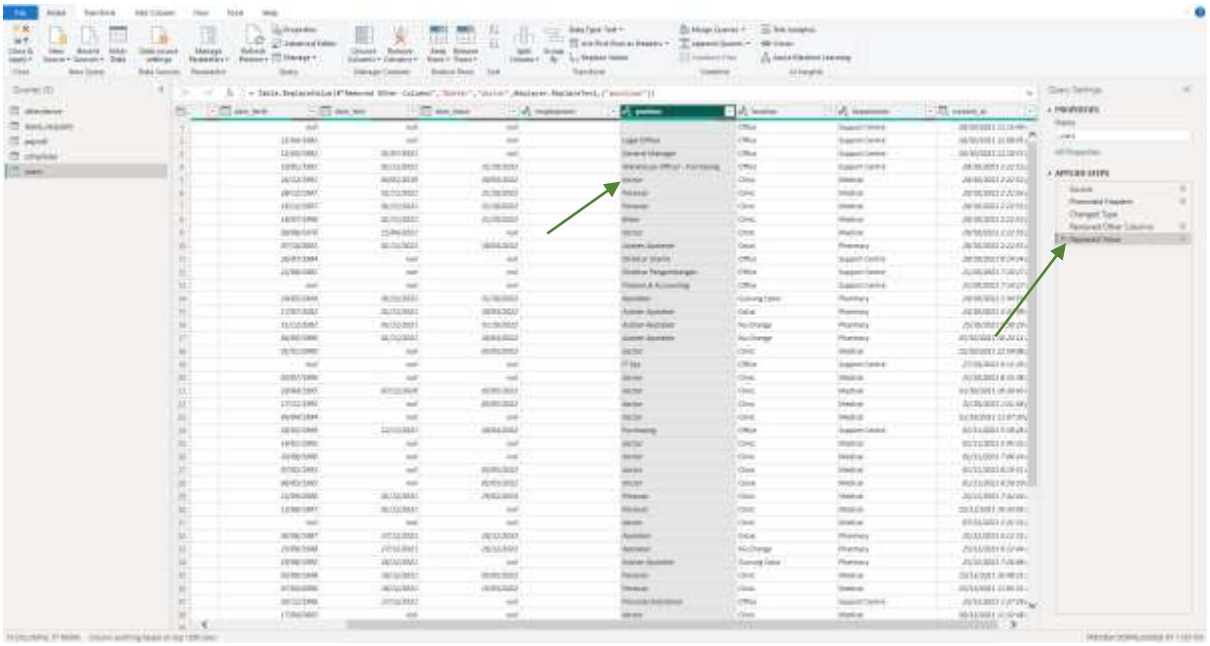
Translate the value of the position column into English. Right click the cell of position column and click replace values.



Input Dokter to Value to Find, in the Replace With input box type doctor the English translation of dokter and then click ok.

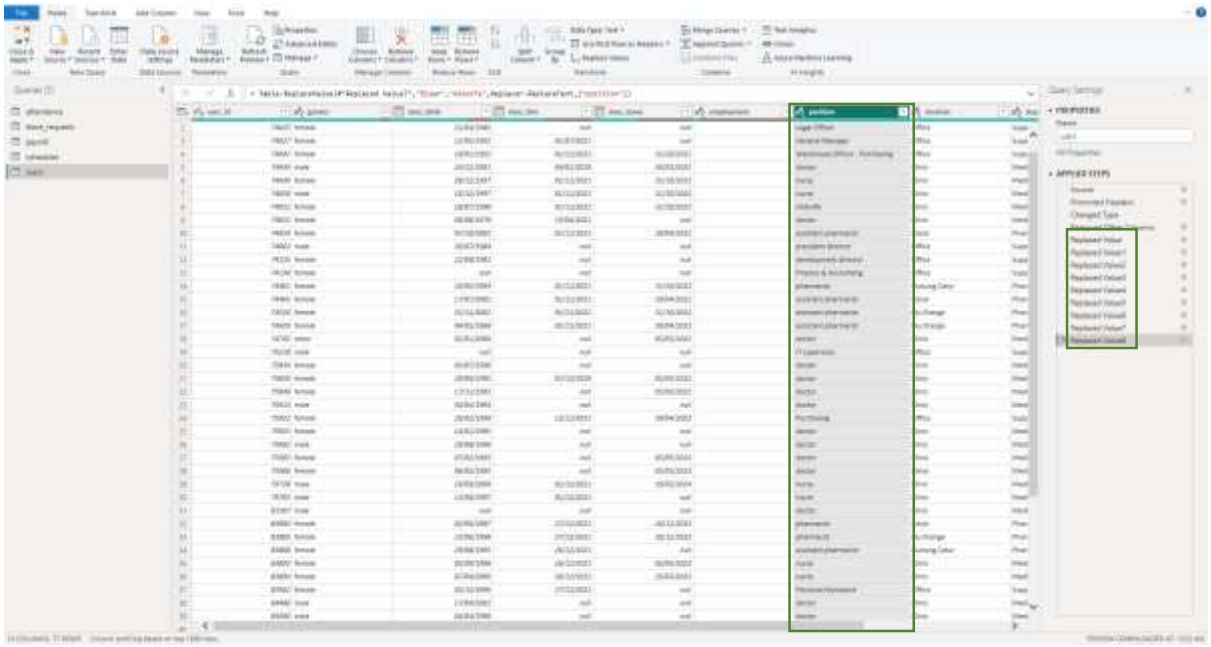


Result of translating the position column values into English.

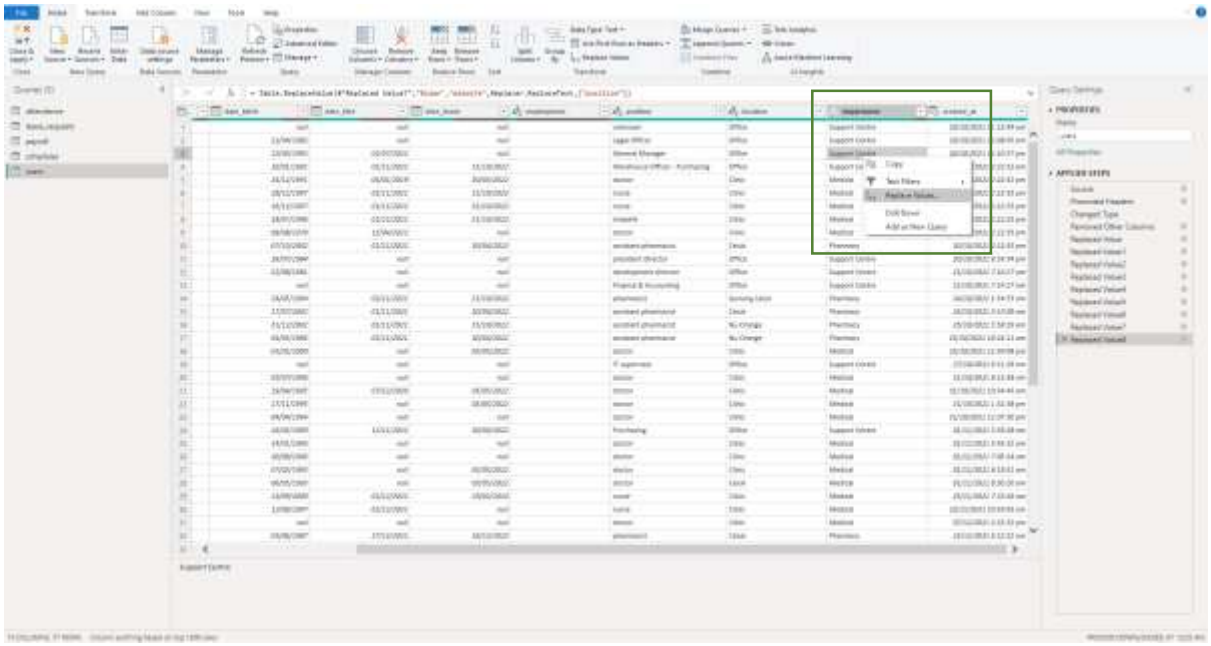


Repeat the same procedure to replace all the position values into English.

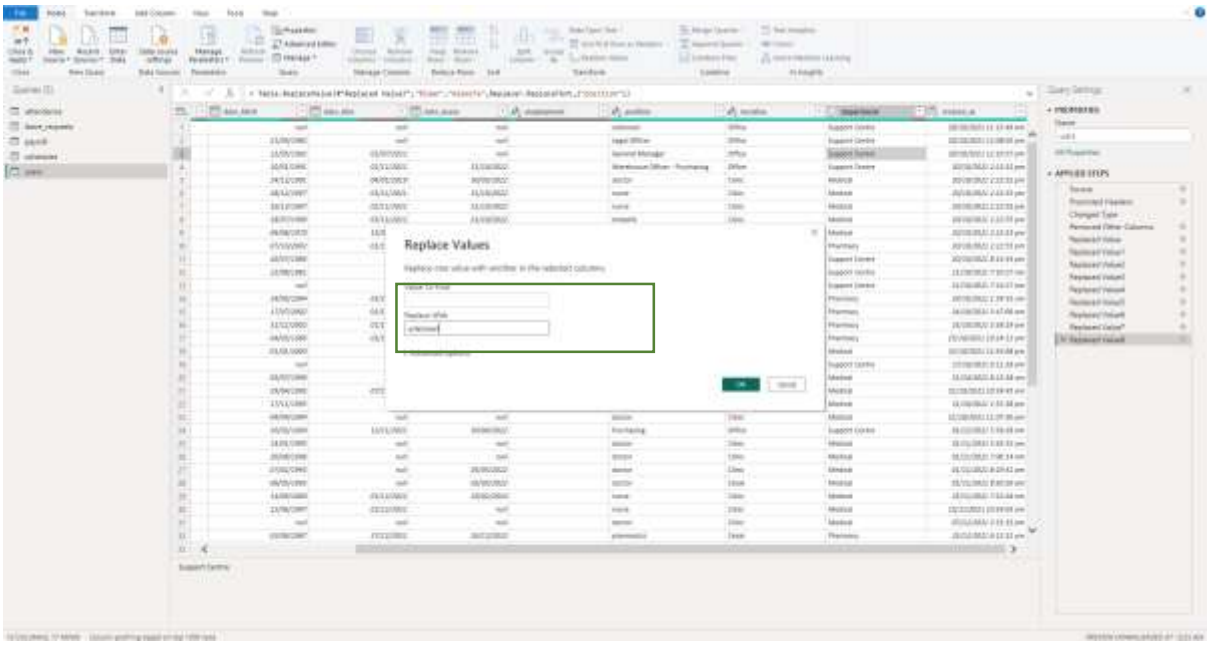
Values to Find	Replace With
Perawat	nurse
Bidan	midwife
Asisten Apoteker	pharmacist assistant
Direktur Utama	president director
Direktur Pengembangan	development director
Apoteker	pharmacist
Admin Bisnis	business admins
IT Spv	IT supervisor
Blank	unknown



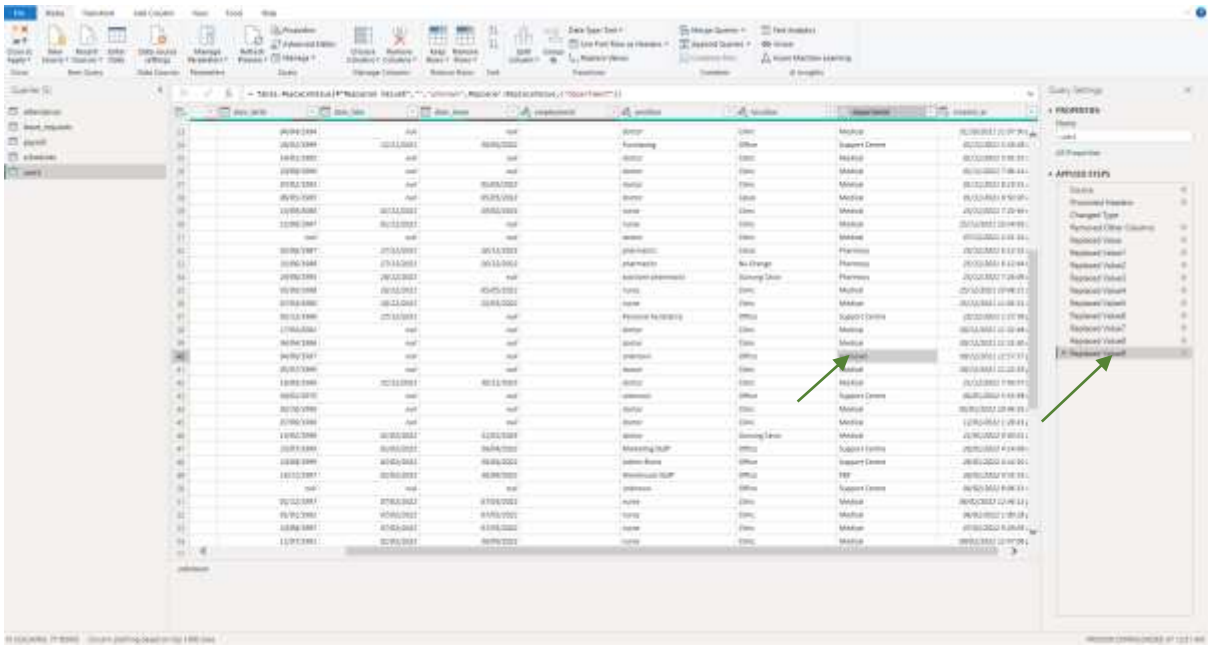
Replacing the blank value of department column. Right click the cell of the department column, click replace values.



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



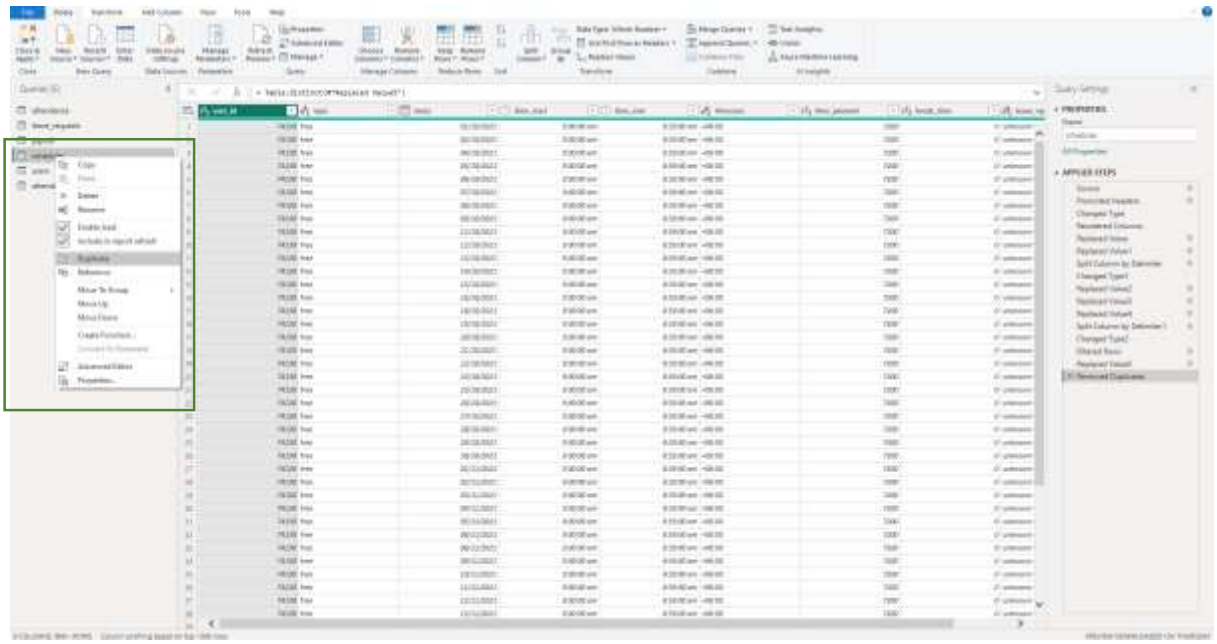
Result of replacing blank value in department column. Click Close to apply to load the data to model.



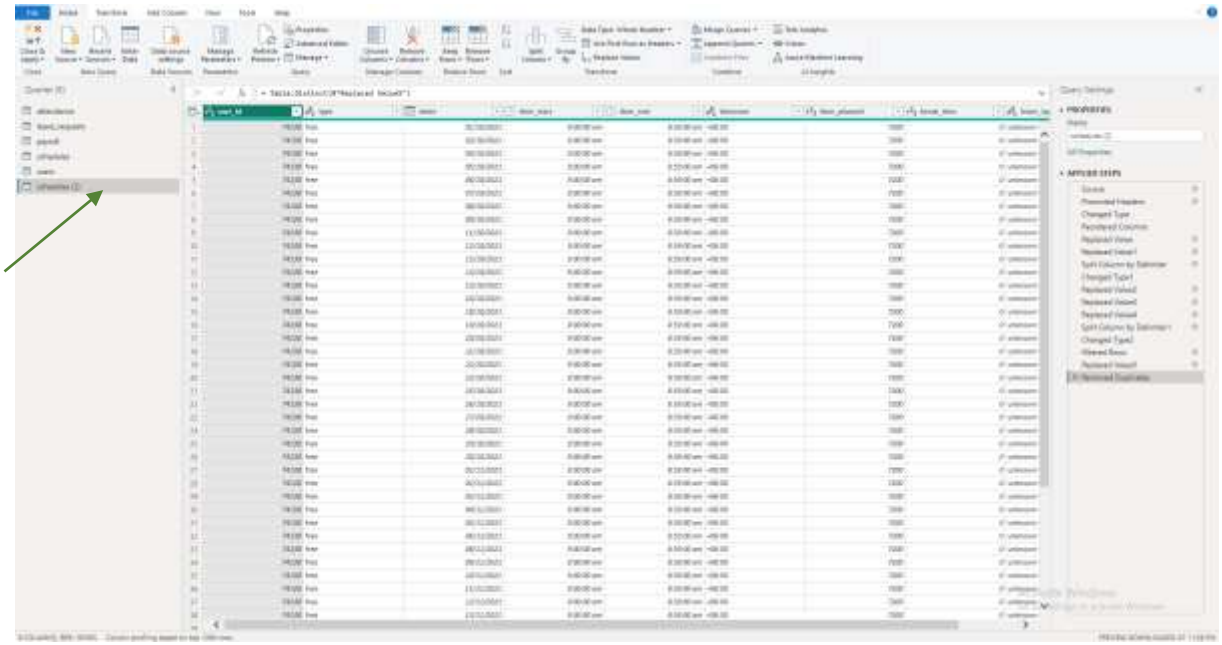
The cleaning and transformation of the five tables is finished. The next step is to merge the schedules table with attendance table.

MERGING ATTENDANCE AND SCHEDULES TABLE

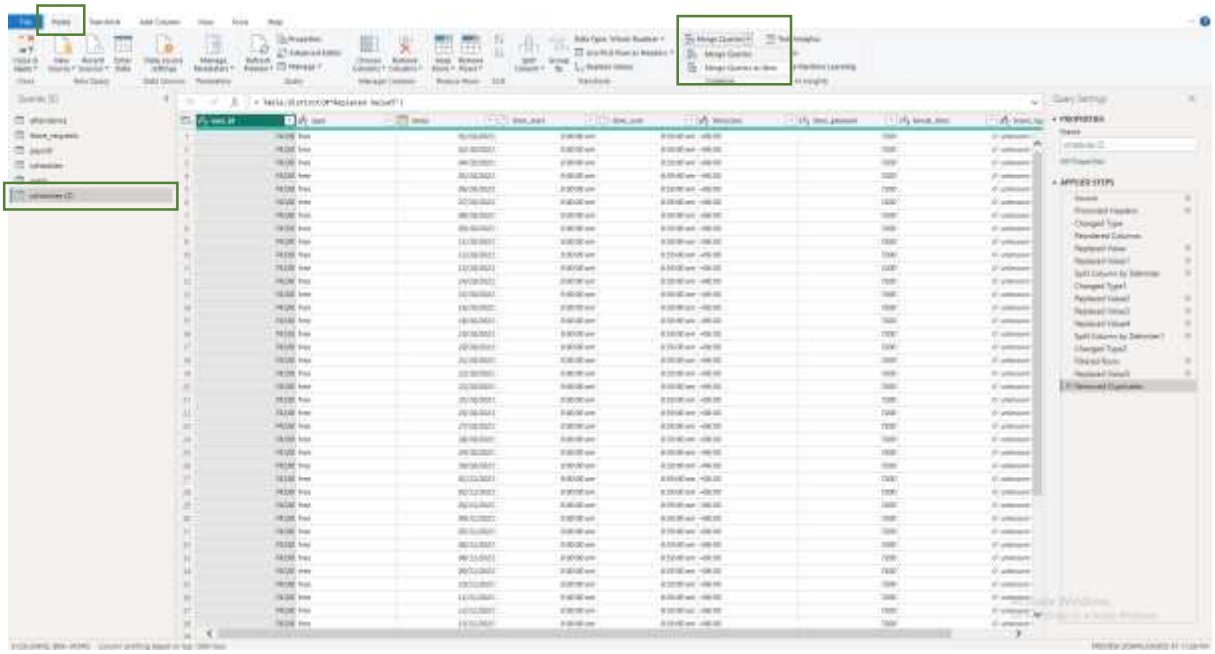
Make a duplicate of the schedules table by Right Clicking the schedules table to the Navigator Pane and click Duplicate.



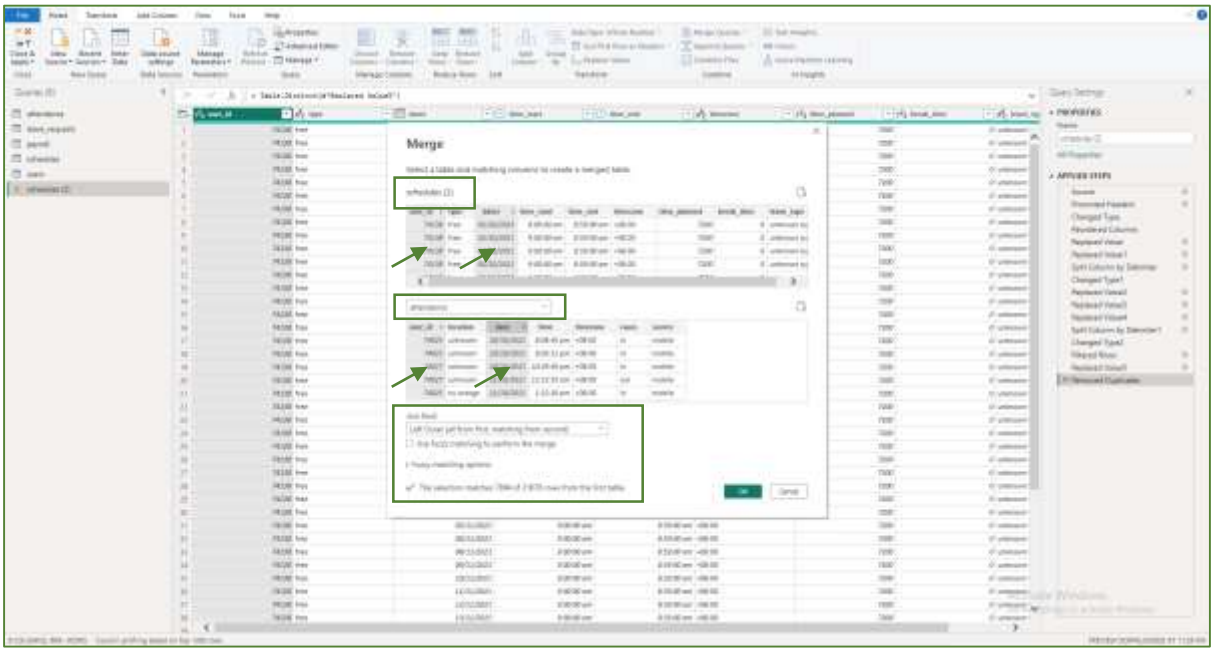
Results of duplicating the schedules table.



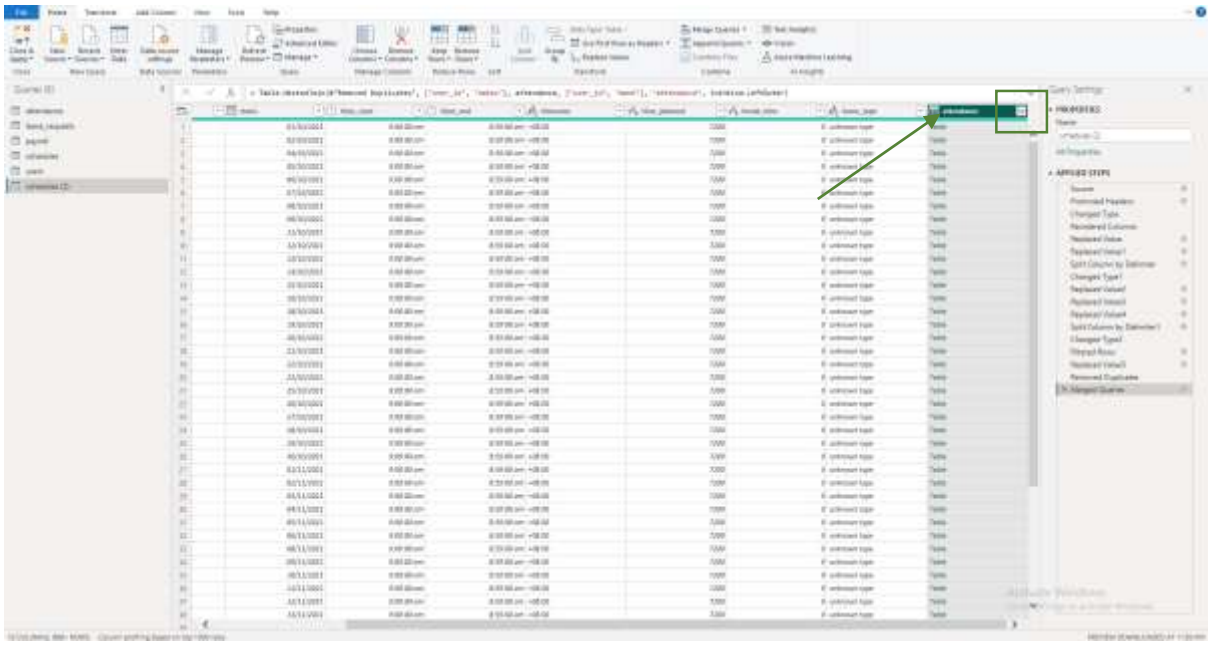
Click the duplicate of schedules table in the Navigator Pane, go to Home and click the drop down arrow of Merge Queries and then click the Merge Queries.



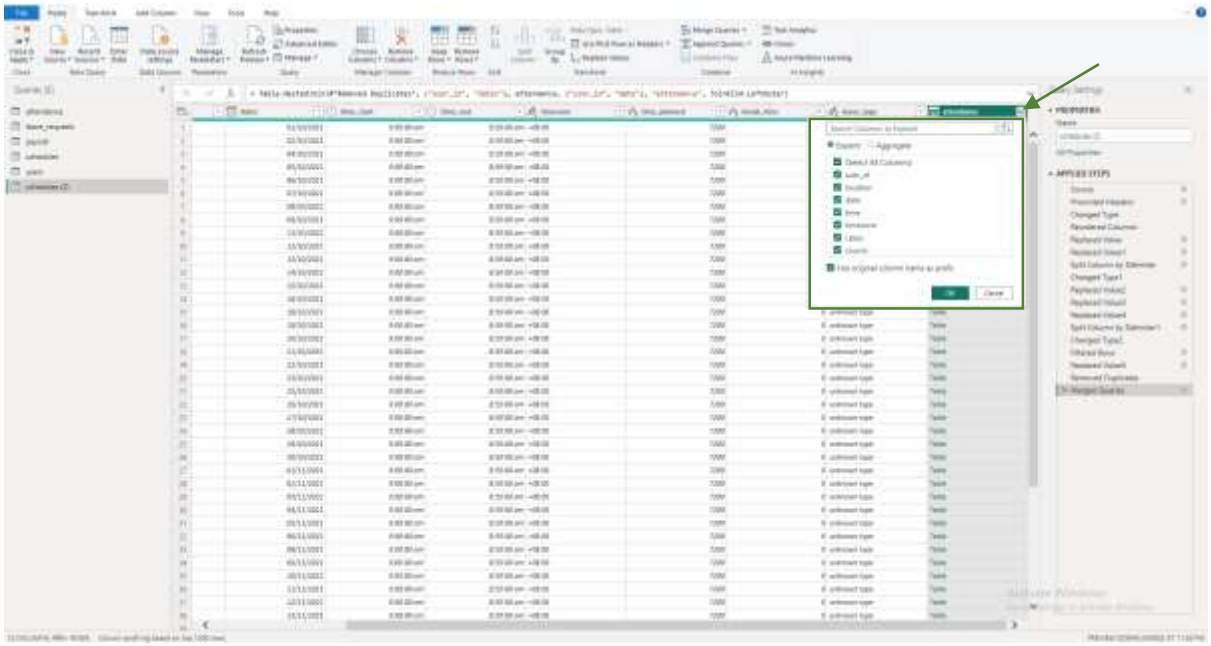
The Merge pop up window appear, this is the format we did when setting the Merge Queries. The first table is the duplicate of the schedules table then the second table is the attendance. The commonality of these two tables is the user_id and dates. Then the Join Kind used is Left Outer. Click ok to continue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



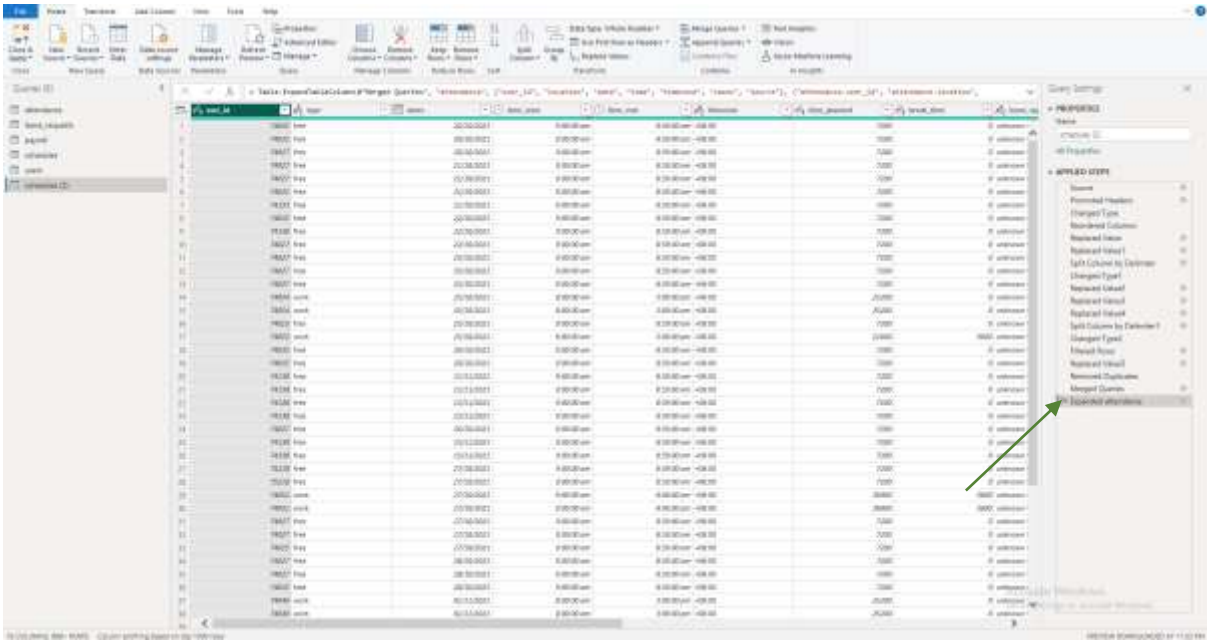
A new column will appear, for now it is still compressed. To display the entire table attendance. Click the icon on the side of the column attendance.



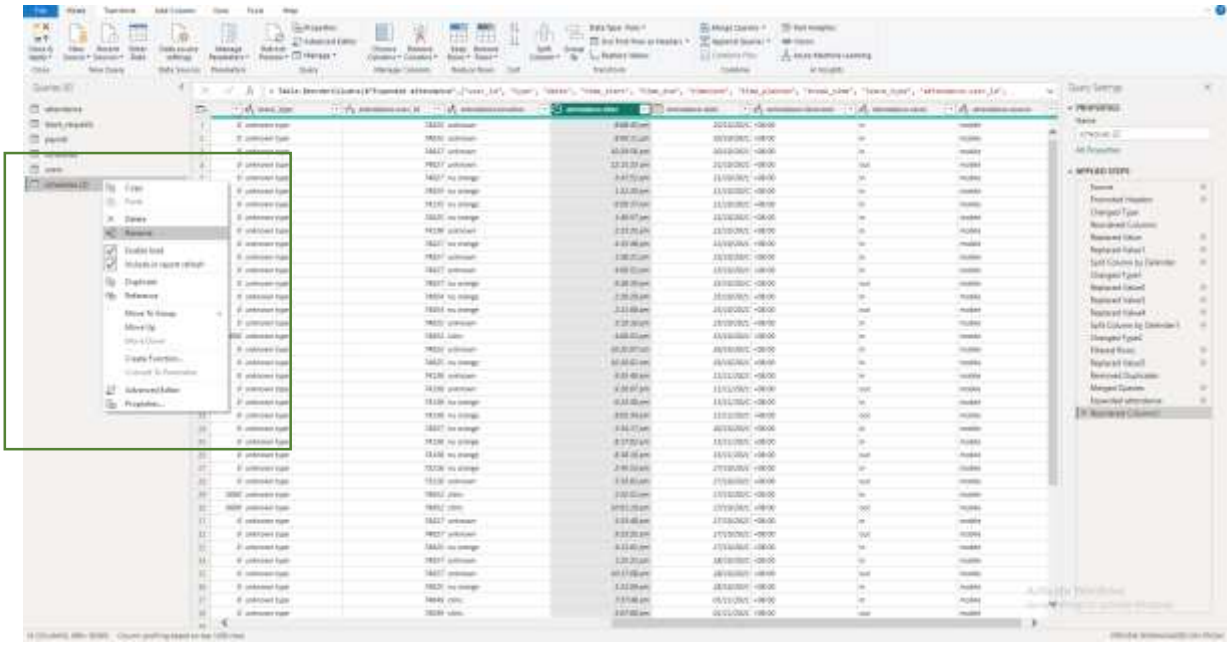
After clicking the icon next to attendance, these check boxes will appear. Here you can select the desired column to appear in the attendance table. Let's leave all columns checked for now.



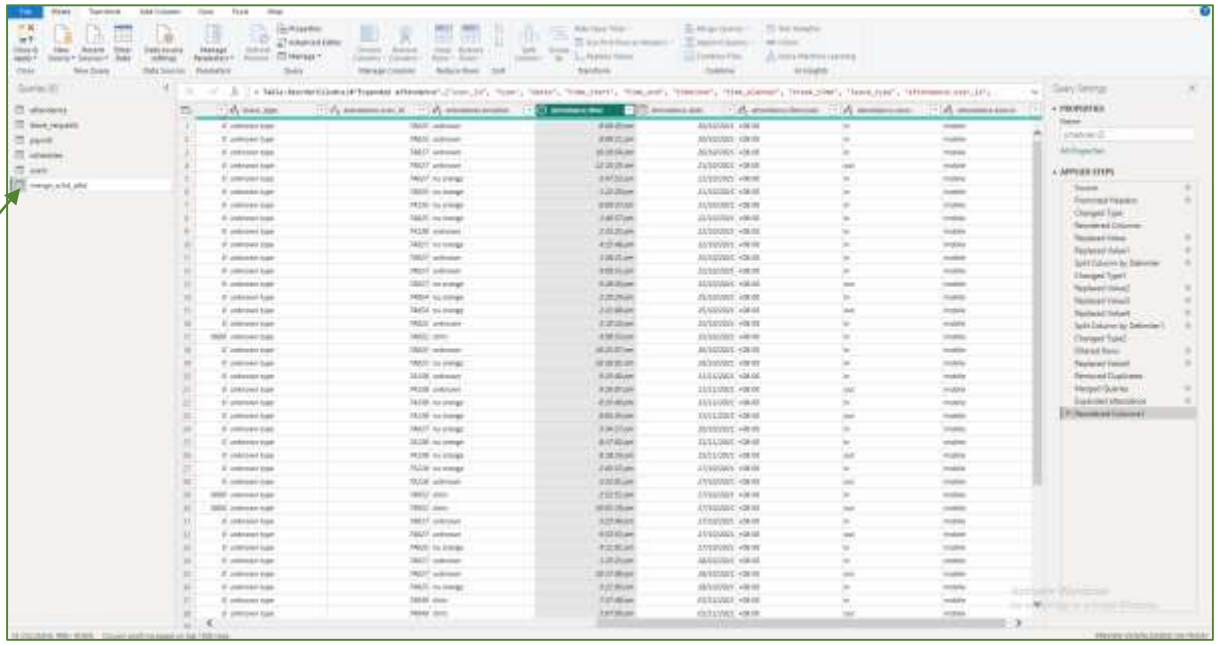
This is the result of merging schedules and attendance table.



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

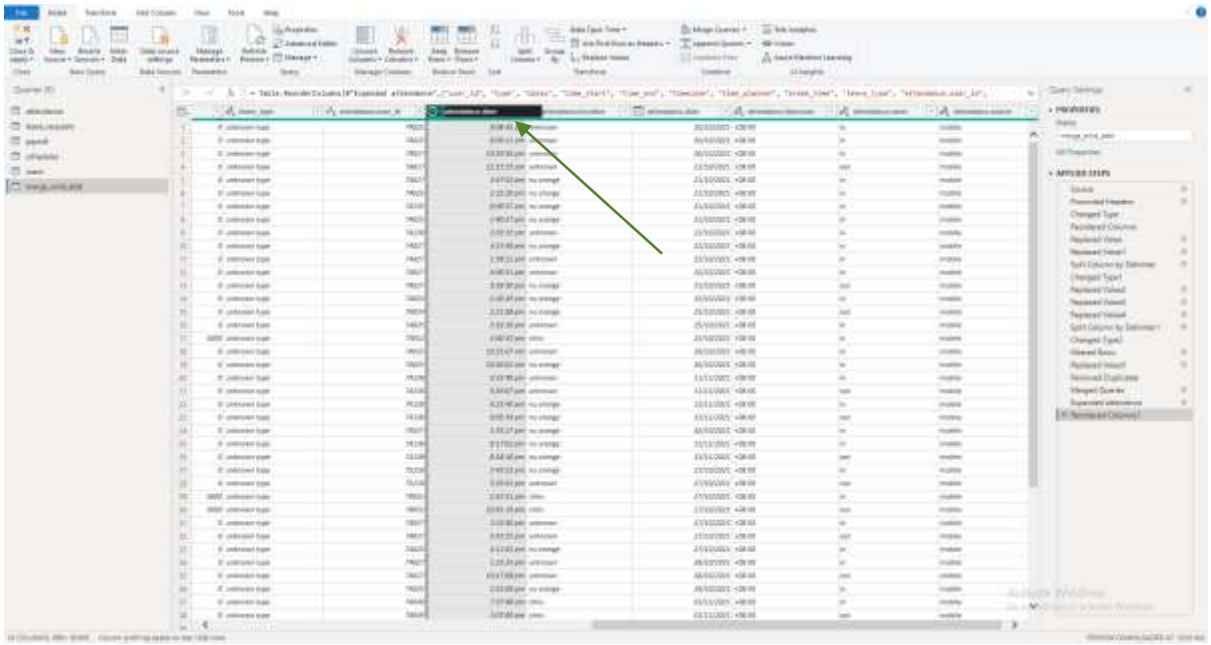


Change the table name to merge_schd_attd.

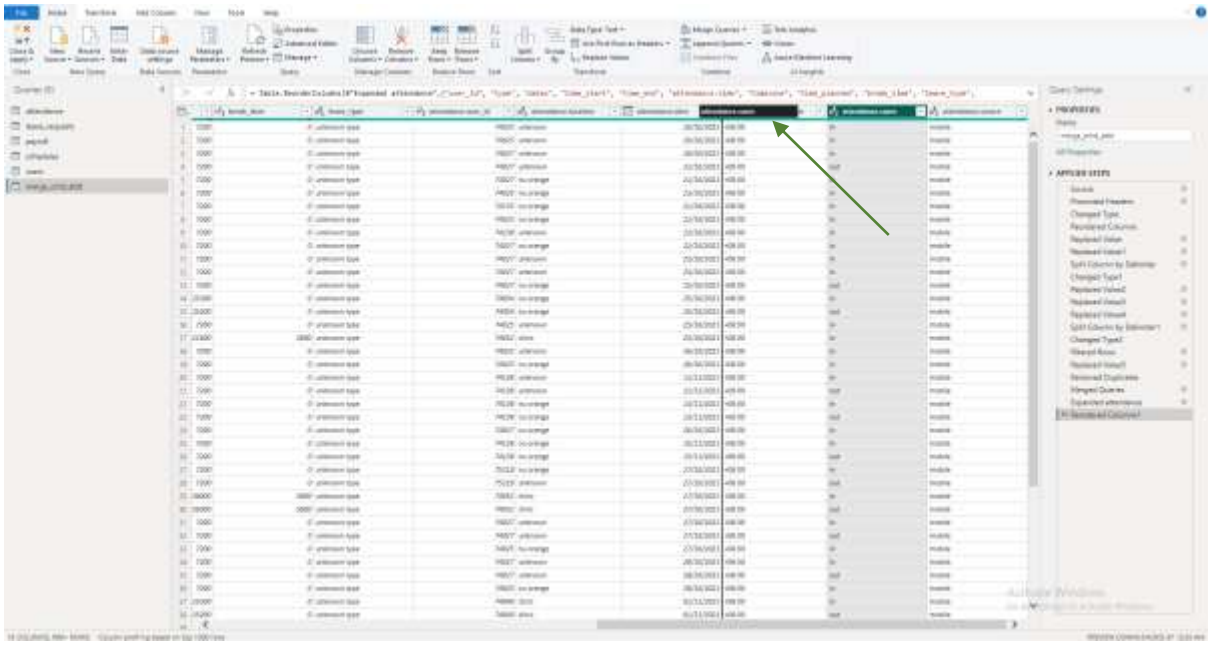


CLEANING AND TRANSFORMING MERGE_SCHD_ATTD

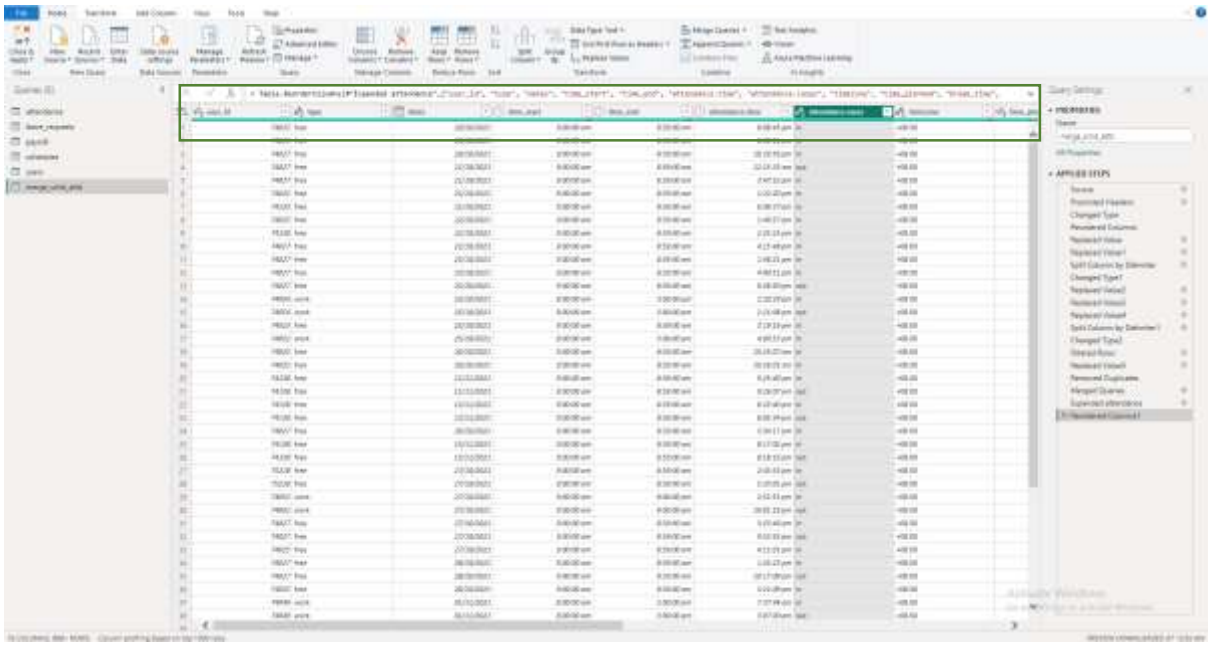
We want to move the position of the attendance.time column and place it next to time_end. Click and hold the attendance.time column and drag it next to time_end.



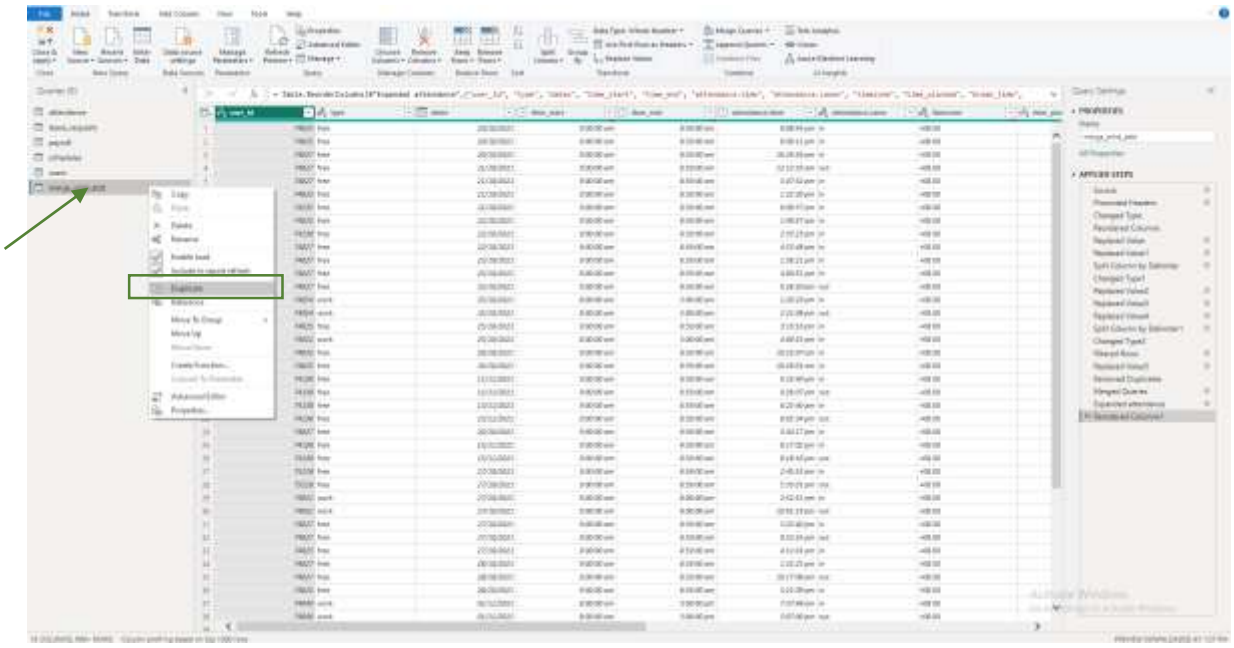
We also want to move the position of the attendance.cases column and put it next to the attendance.time column. Click and hold the attendance.cases column and then drag it next to the attendance.time.



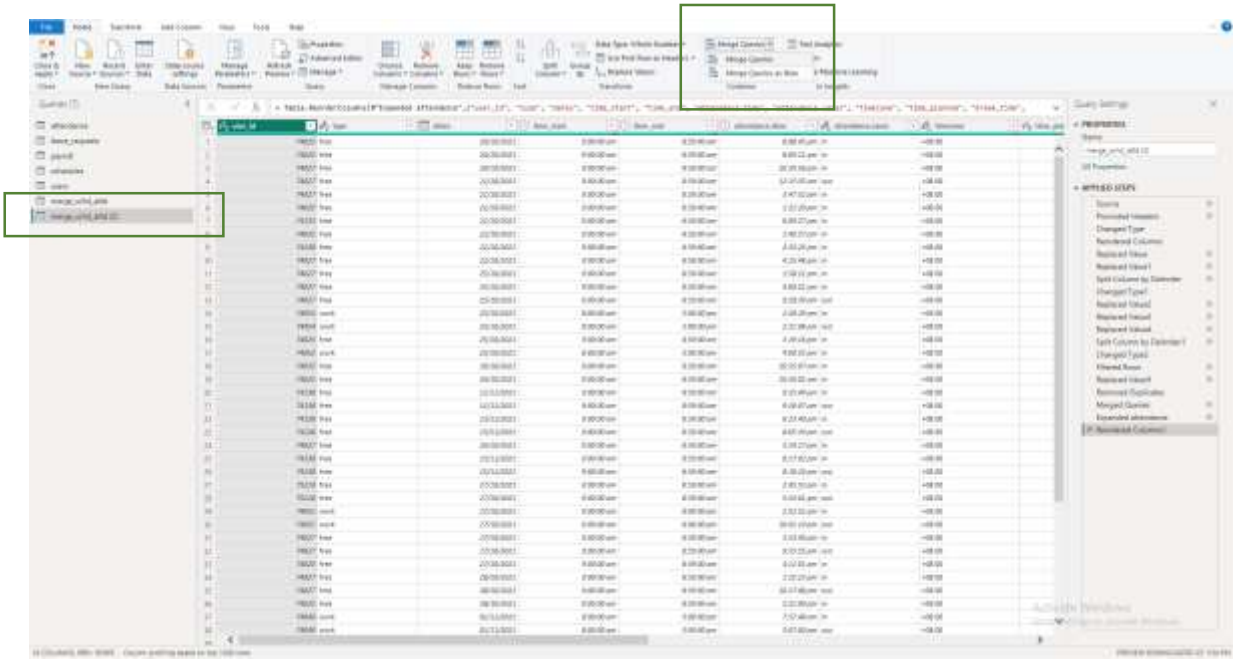
This is the result of moving to the position of attendance.time and attendance.cases



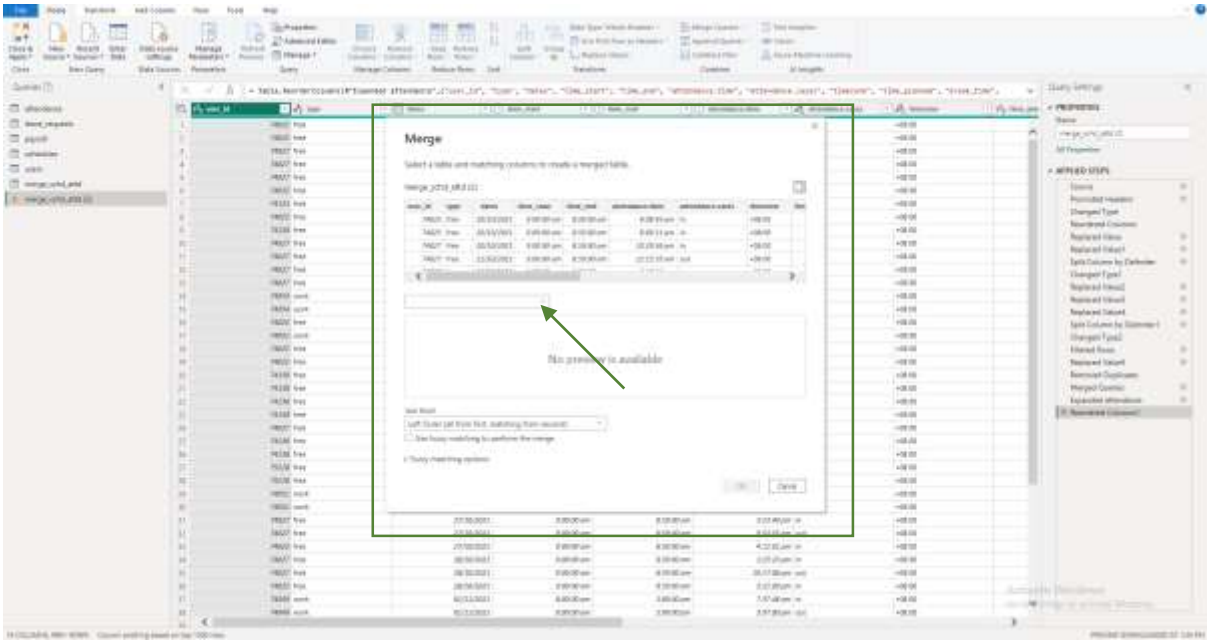
To arrange the data in one row, this is what we did,duplicate the merge_schd_attd table.

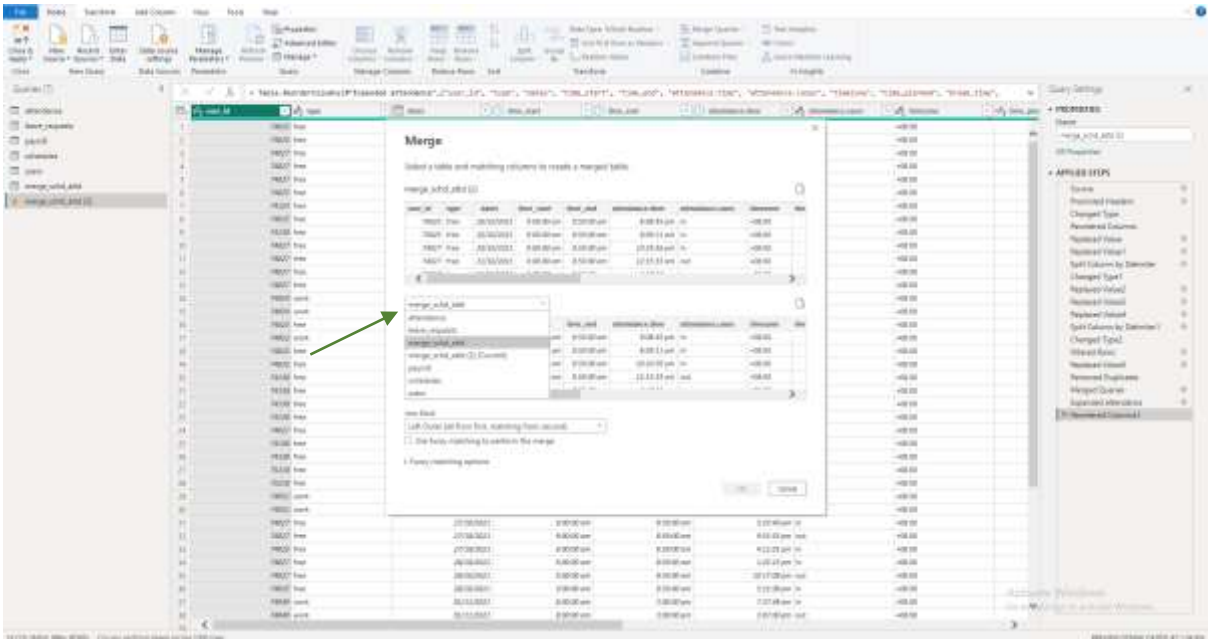


Now that we have duplicated the merge_schd_attd table, let's_merge merge_schd_attd with merge_schd_attd (2). Go to Home, press the Merge Queries drop down arrow.



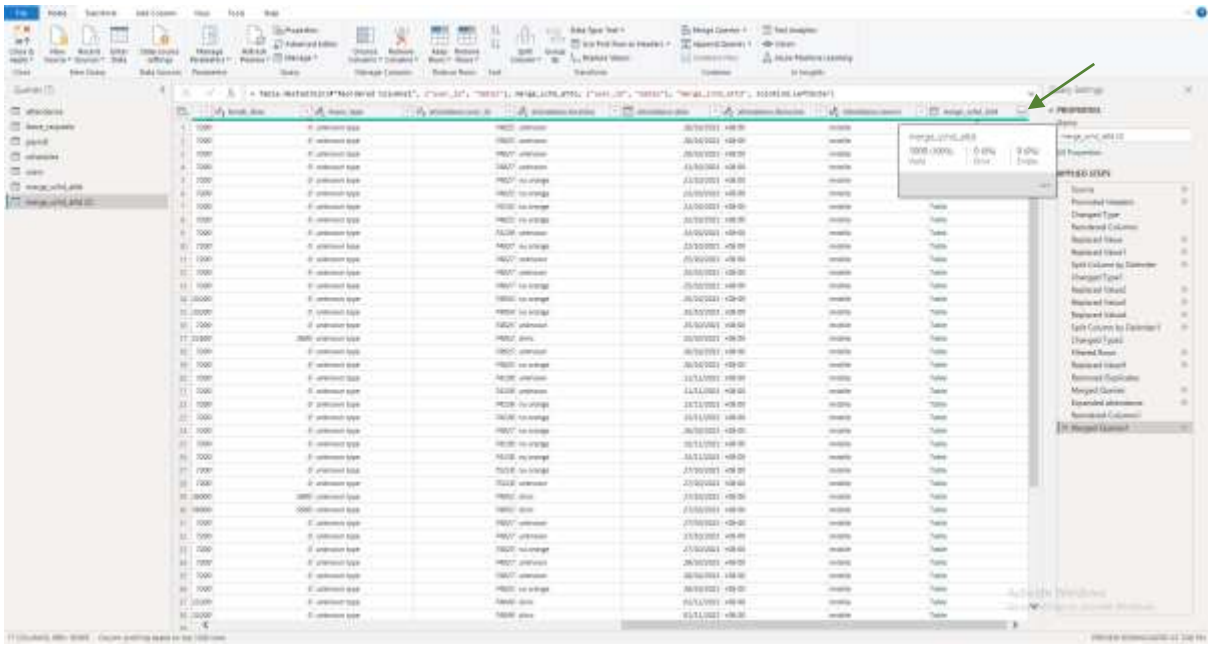
The Merge pop up window will appear. Press the drop down arrow to find the table named merge_schd_attd.



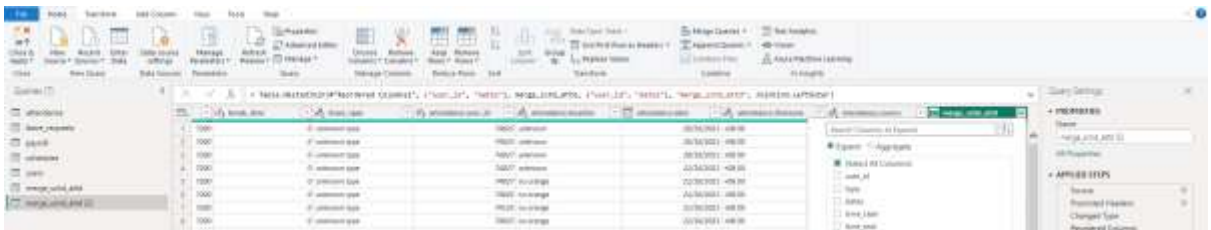


Set the commonality of the two tables using the user_id column and dates.Join Kind is Full Outer, and press ok.

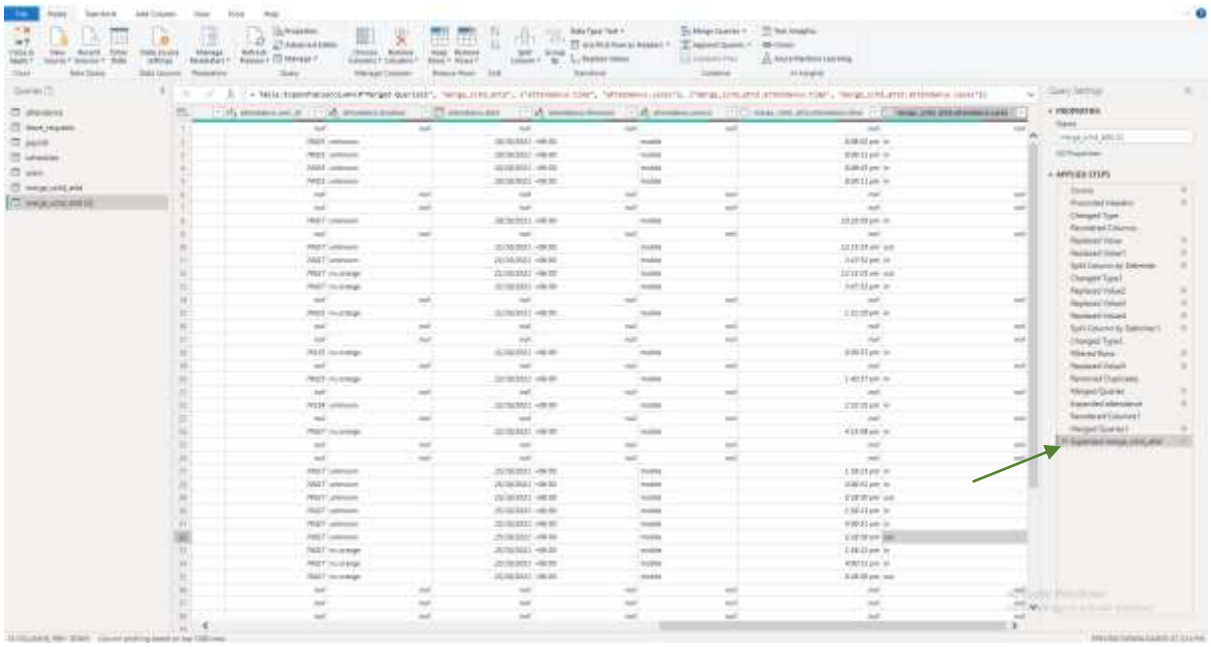
The new merge_schd_attd column will appear. We have to press the icon next to the column name to expand and show the result of the merging process that we did in the two tables.



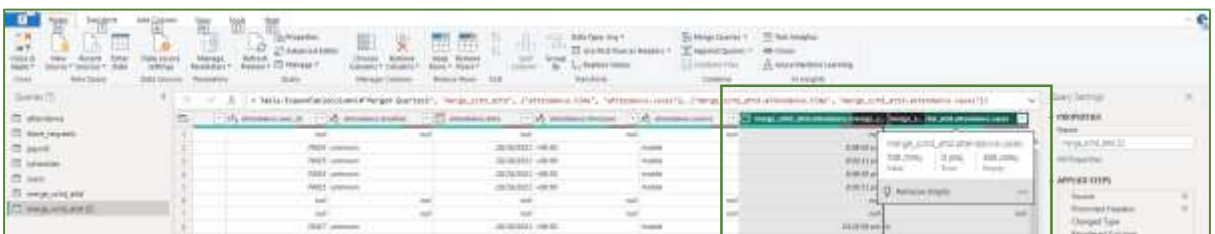
Leave only the attendance.time and attendance.cases column checked and click ok.



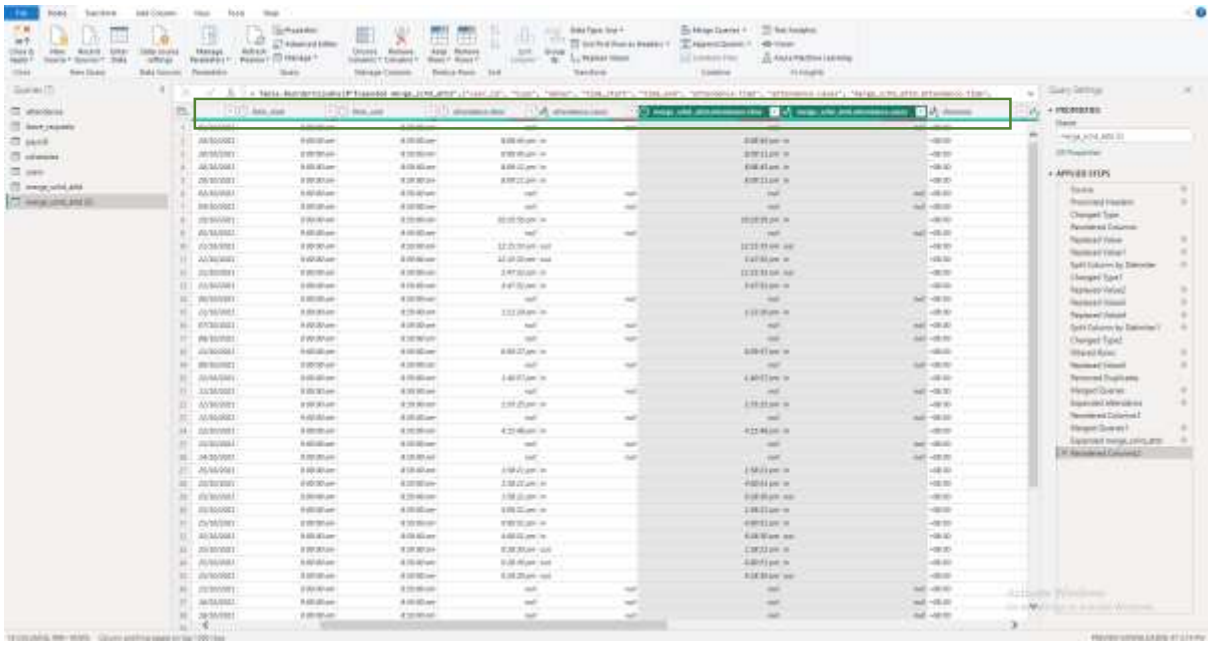
The result is 77641 total rows after merging merge_schd_attd and its duplicates.



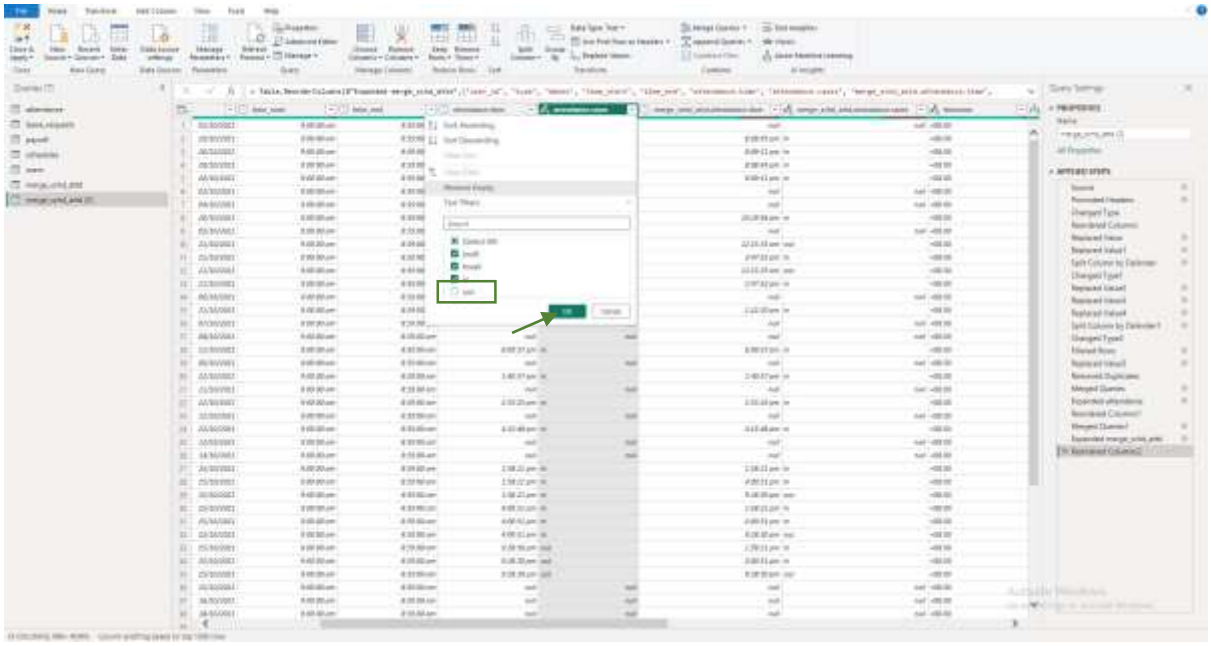
Change the position of the new column next to attendance.cases. Highlight the two columns and click and hold then drag it to where you want to put it. In this case next to the attendance.cases column.



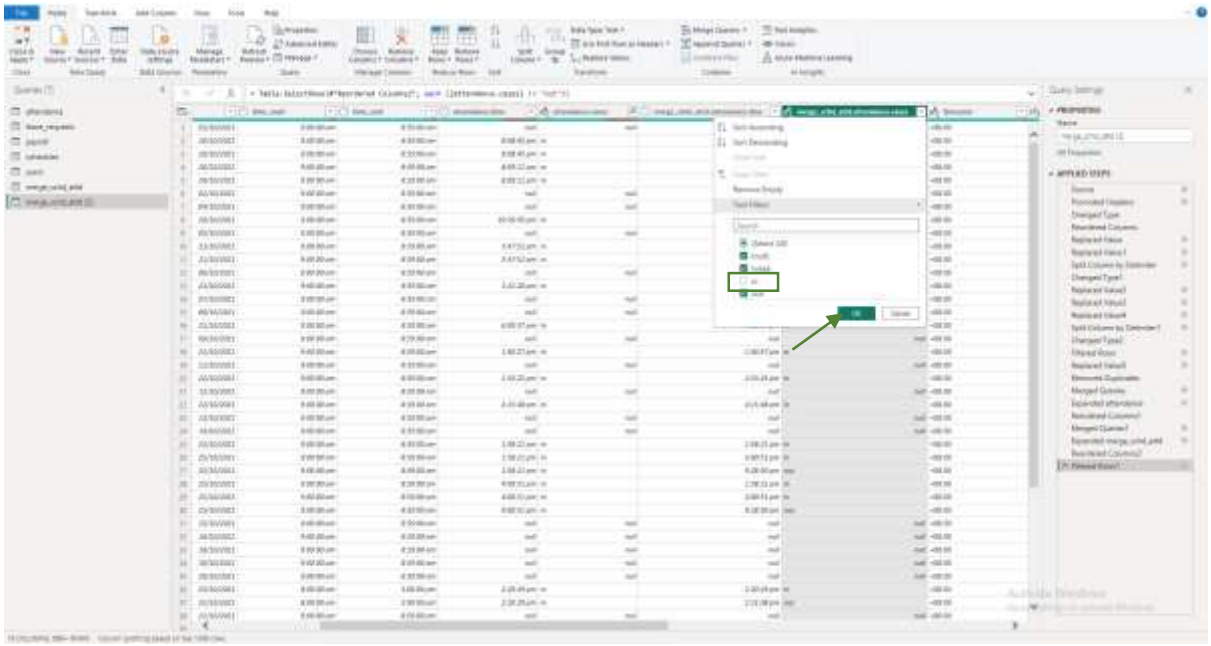
This is the result, the two new columns are next to attendance.cases.



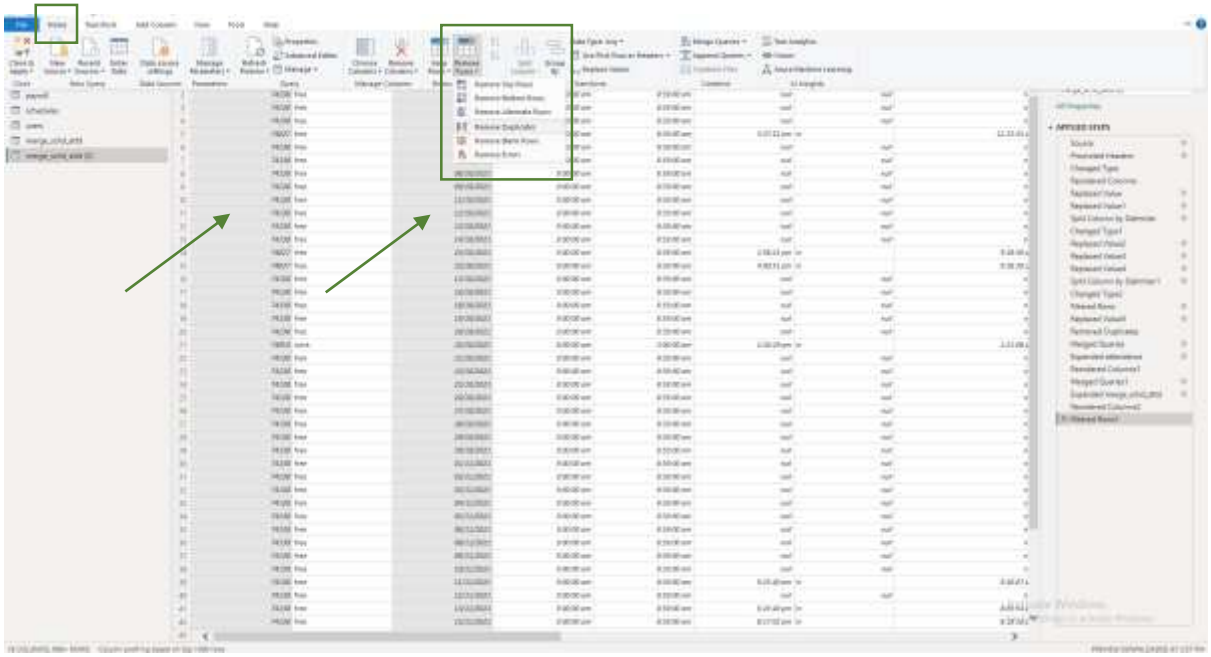
In attendance.cases press the drop down arrow next to the column name. uncheck the “out” value and then click ok.



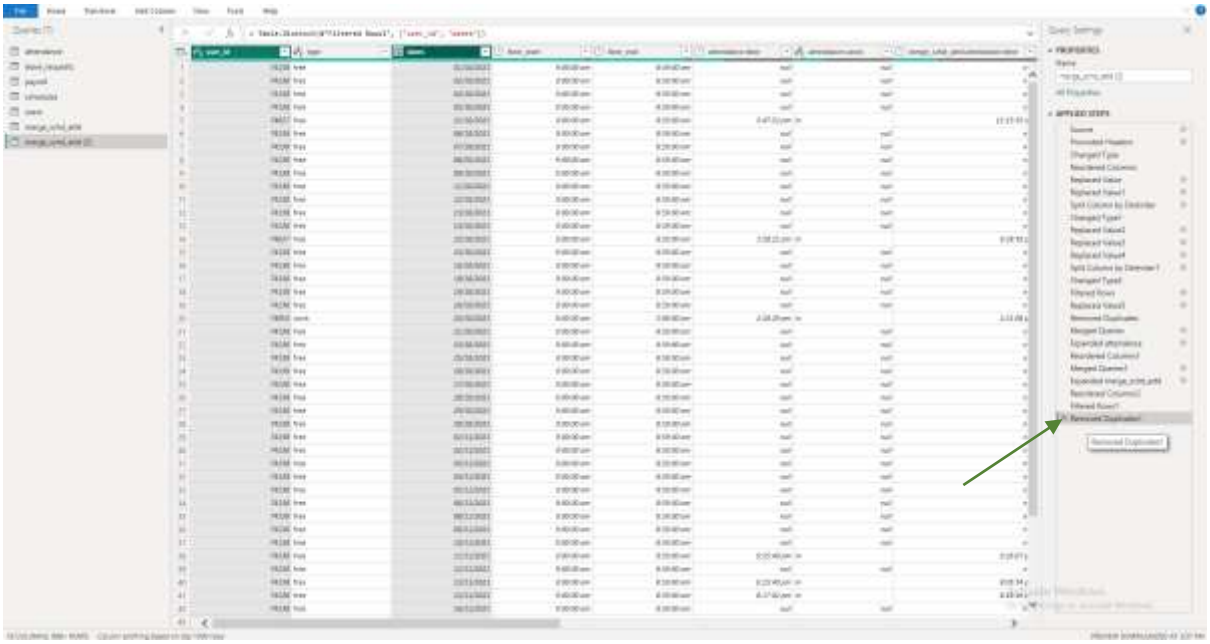
In merge_schd_attd.attendance.cases press the drop down arrow next to the column name. uncheck the “in” value and then click ok.



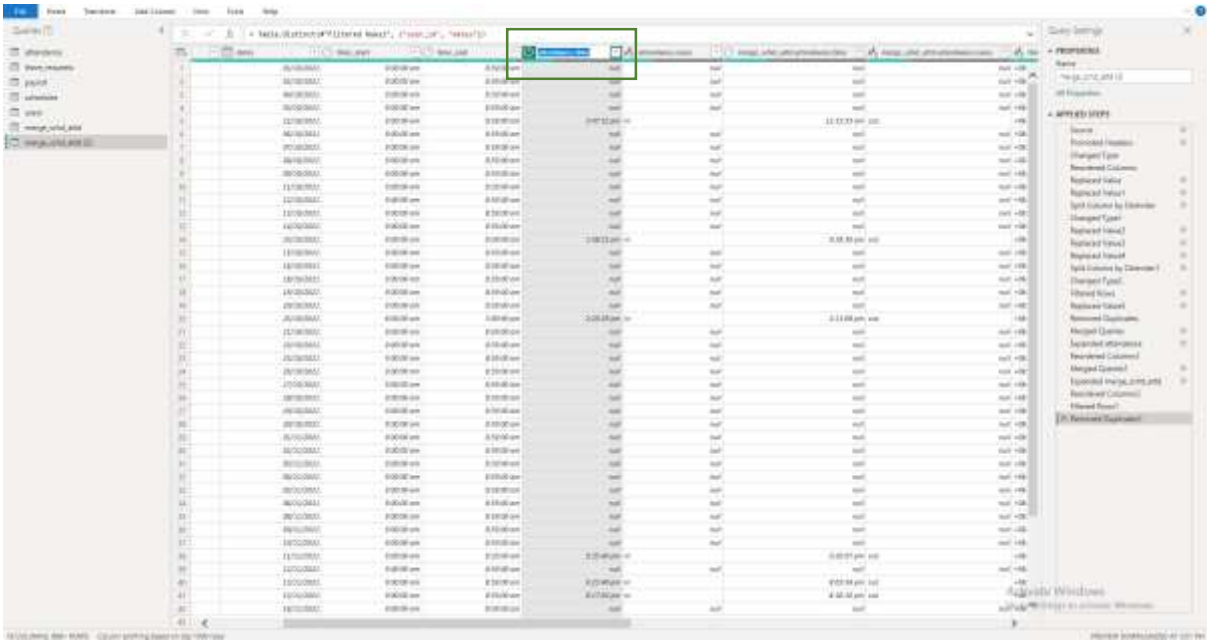
Let's remove the duplicates in this data table. High light the column of user_id and dates. Go to Home, press the drop down arrow of Remove Rows then click Remove Duplicates.



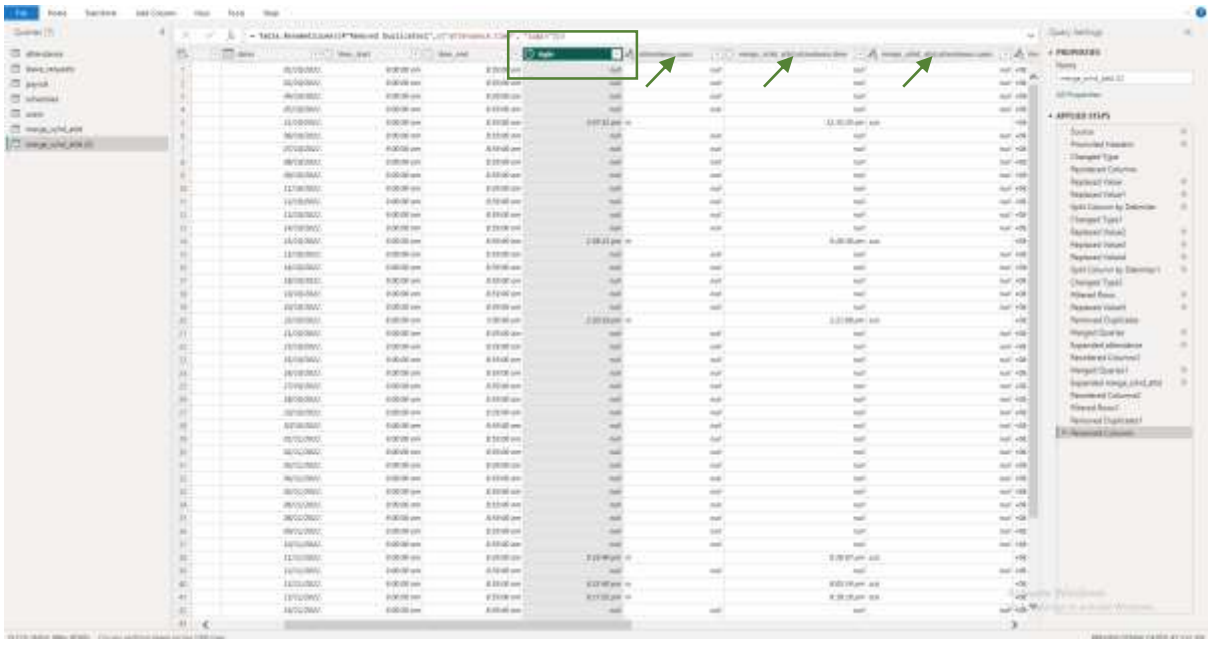
The total rows of our final table are 21247.



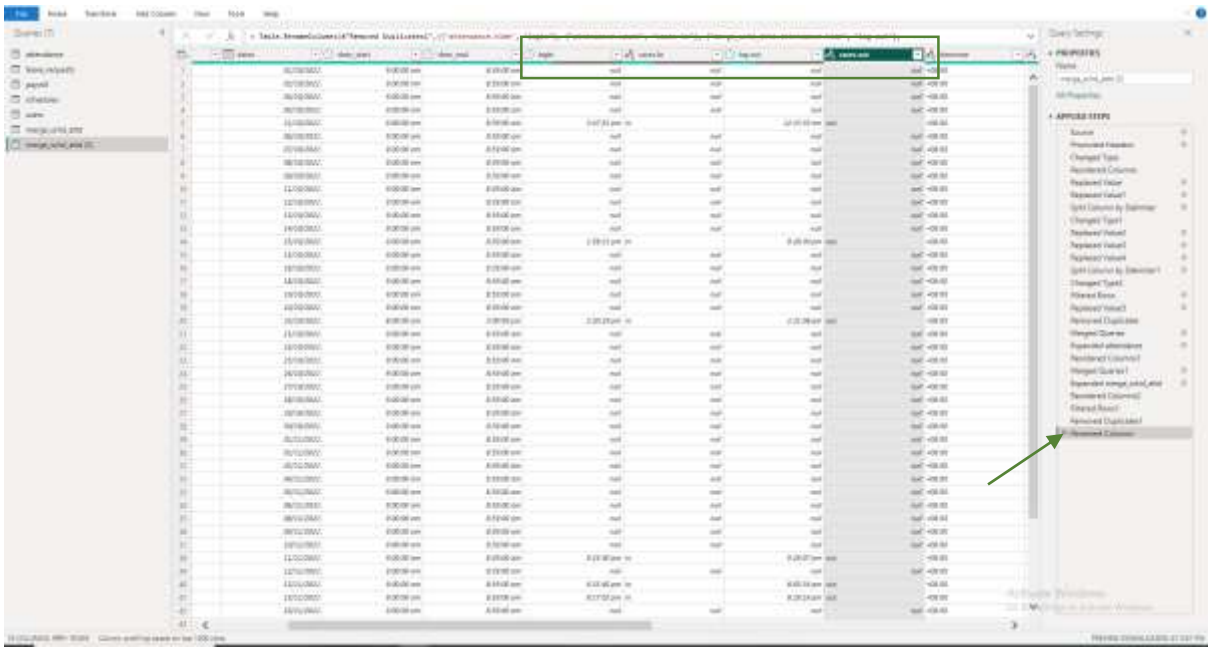
Let's change the column name attendance.time to “login”. To do this, double click on the column name to highlight it, then type "login" this will be the new column name



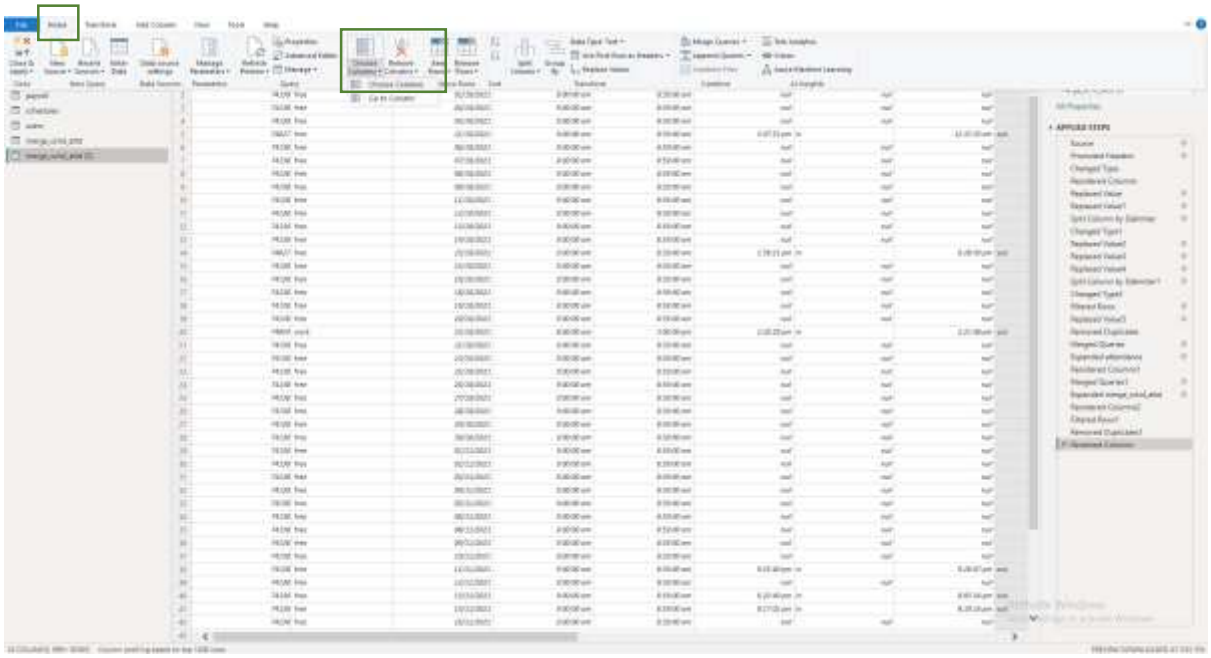
The previous attendance.time column name has become "login". Repeat this process to replace the attendance.cases column with "cases-in", merge_schd_attd.attendance.time with "log out", merge_schd_attd.attendance.cases with "case-out". This is to make it look better and easier to understand.



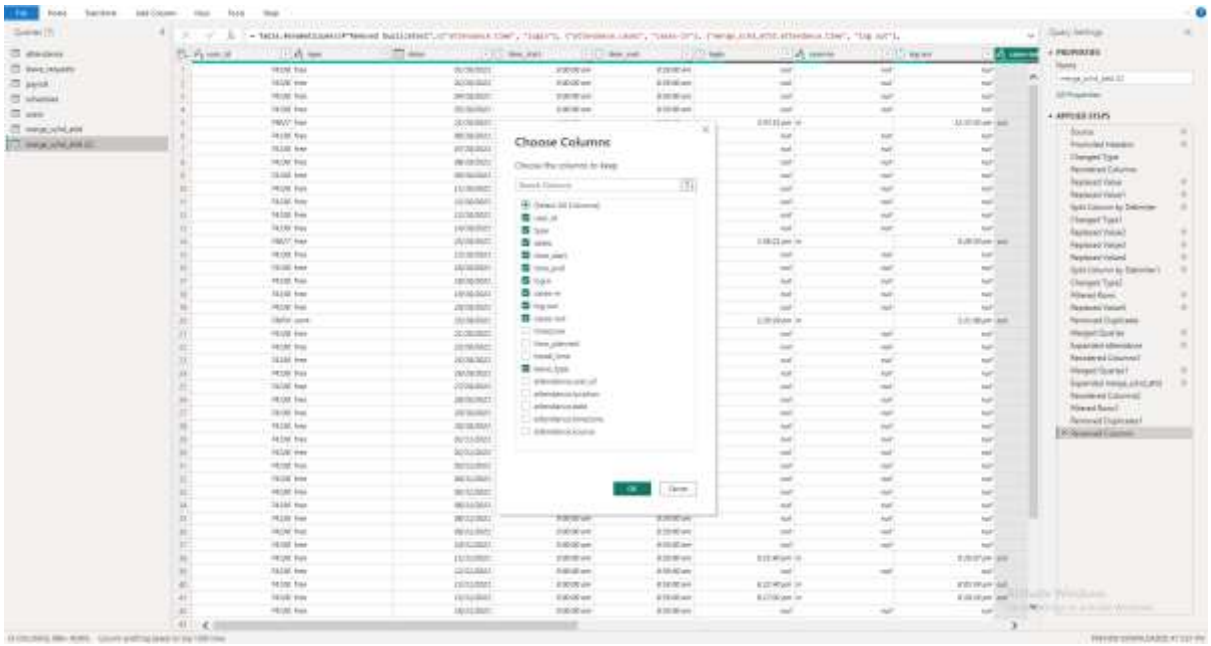
We have changed the column name of four columns.



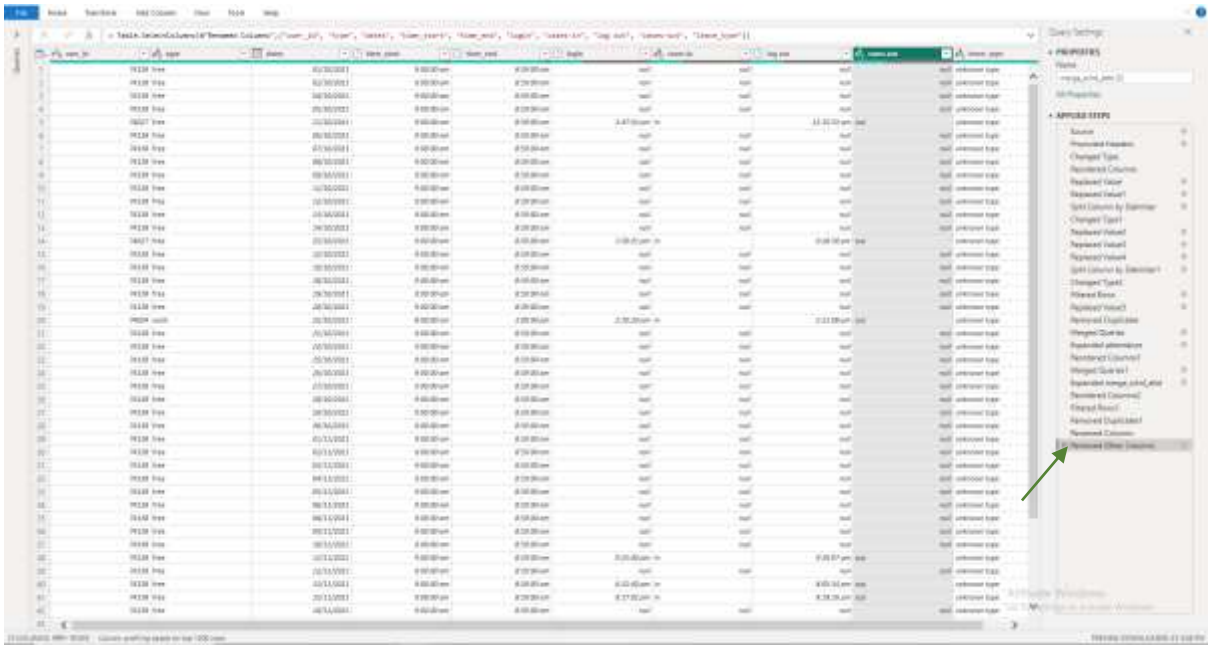
Let's delete other columns that are not needed in our table. Go to Home, press the drop down arrow of Choose Column and click Choose Column.



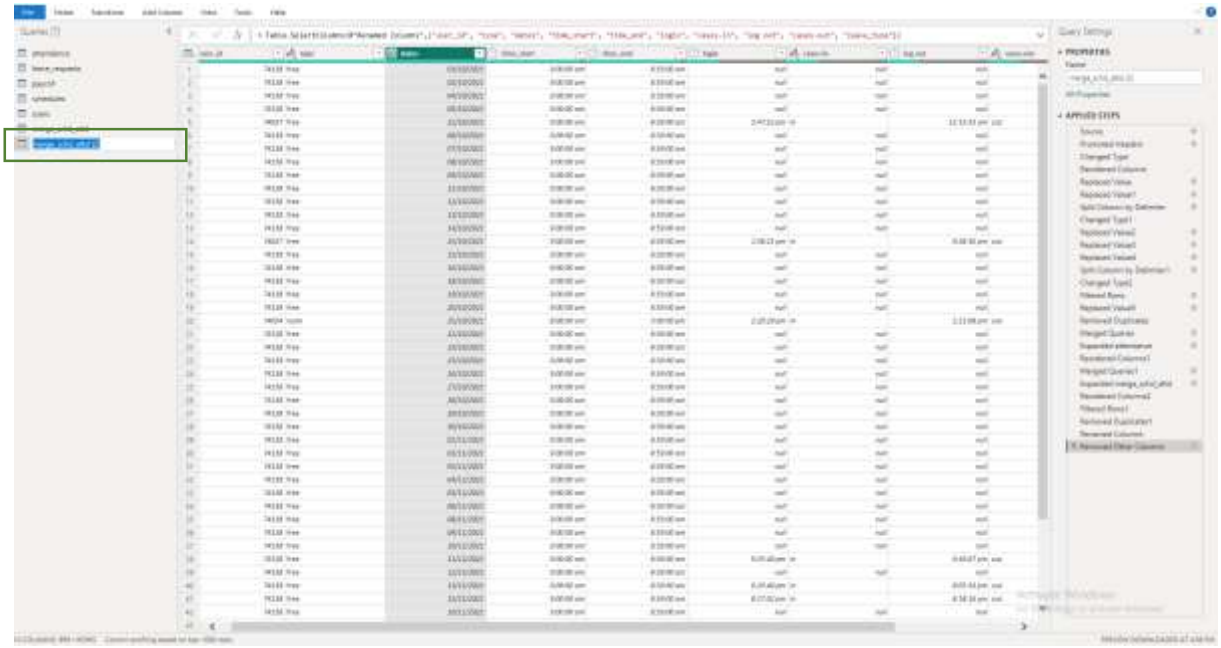
Leave the checkbox of user_id,type,dates,time_start,time_end,login,cases-in,log out,cases-out and leave_type checked and uncheck the remaining column to remove it from this table. Click ok to continue.



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



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.



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